High-k Dielectric Stacks for Integration into an Advanced CMOS Process

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

by

William Mark Davey

March 2010

Contents

Abstract

This thesis details the work carried out to analyse dielectrics for possible use as replacement gate materials for silicon dioxide in conventional CMOS transistors. The four year project analysed many materials, however two promising candidates were identified and further analysis and development was carried out. The first material was amorphous hafnium silicate, $(HfO_2)_x(SiO_2)_{1-x}$. The electronic k-value and the optical dielectric functions were both seen to increase with hafnium content, however, as expected the band-gap of the film decreased. Trapping levels were observed using spectroscopic ellipsometry (SE) and as previously reported occurred deeper with increasing silicon content up to 50%. A density estimation technique was proposed using the Clausius-Mosotti equation and the change of density was studied with composition. The accuracy was within 15% that of a standard technique using x-ray reflectometry. SE, medium energy ion scattering (MEIS) and transmission electron microscopy (TEM) were also compared as a thickness measurement technique.

The second material studied in depth was gadolinium oxide. Reducing the initial interfacial layer (IL) was found to decrease the capacitive equivalent thickness (CET) of the stack after rapid thermal anneal (RTA) for thick layers. However, residual oxygen during the RTA process limited the CET reduction for thin initial IL. An RTA procedure was developed to avoid the CET limiting IL re-growth and the requirements for the 45nm node for low standby power were achieved within a manufacturable process. A study was made to match the performance of the e-beam evaporated layers using Atomic Layer Deposited (ALD) silicated gadolinium oxide films. Silicon incorporation was realised during deposition and also by interfacial layer consumption during RTA. For silicon incorporation during deposition, the highest permittivities were seen for the lowest deposition temperatures and the band-gap for thinner samples was close to that measured for epitaxial gadolinium oxide. For the gadolinium silicate layers formed using RTA the silicon incorporation was observed to reverse the crystallisation of the gadolinium oxide at high temperatures. The ALD layers were unable to meet the ITRS 45nm node requirements due to both precursor growth rate instability and due to the same residual oxygen problem seen in the e-beam evaporated samples.

Acknowledgements

I would first like to thank my supervisors for their support and guidance during my PhD course. Even though Professor Stephen Hall was kept very busy for most of the last four years with departmental administration he was always able to find time to assist and encourage me with work and for that I am very grateful. Even though he had to unfortunately leave halfway through my studies I would also like to also thank Dr Octavian Buiu. His cheerfulness and humour always got me through the most frustrating of days. I am also indebted for his demonstration of ellipsometric techniques and for his encouragement in both my work and home life. I am also thankful to Professor Hall's secretary Jacqueline Cowan for always providing a happy face when I came to his office and always lending an ear to listen to any problems I had.

I would like to thank Dr Matt Werner for the long hours he worked both growing the ALD gadolinium oxide films and performing the MEIS measurements on them. I will long remember the chats we had late in the evenings playing 'Russian roulette' choosing ALD cycle numbers due to the problematic precursor growth rates. On a similar note I am also very grateful to my group colleagues Dr Ivona Mitrovic and Dr Lu Yi for their assistance with electrical techniques and useful suggestions during data analysis

I would like to thank Heiner, Matthias, Andrea and my other colleagues at AMO AMICA also for always welcoming me to Aachen and for allowing me the use of their laboratory. I am also grateful to them for their teachings and demonstrations on the production side of semiconductor research.

I also would like to thank Dr Robert Tan, with who I have now shared an office with for 4 years, for his fruitful suggestions on work, for listening when work frustrated me beyond my limits and for providing me an outlet for my dry sense of humour.

I would also like to acknowledge my many friends I had in the department including James, Vivek, Robert, Tom and many others I haven't got the space to mention. I can safely say it wouldn't have been as fun without you all.

I would also like to acknowledge the financial support of the EPSRC and the European projects PULLNANO and NANOSIL for their financial support of both the research carried out and for providing funding for my PhD course. I am also grateful to IMEC for supplying the hafnium based samples for this work and for carrying out XRR measurements to verify the density estimation technique.

I thank my parents for their love, for giving me such a good upbringing and for the support they have always given me. For his belief in me and support through whatever choices I made I would also like to thank my granddad and I only wish I had been able to tell you how much it meant to me.

Finally, I couldn't finish an acknowledgement section without mentioning my fiancé Lindsey. I would like to thank you for not only the numerous cups of coffee that you made me through the long nights and for your unshakable confidence in me, but also for the many fun times you have given me. Hopefully I'll be able to bore you by explaining the subtleties of solid state electronics (whether you want to know or not) for many years to come.

William Davey, March 2010

Glossary of abbreviations

ALD	Atomic layer deposition
CET	Capacitance equivalent thickness
C-V	Capacitance-voltage measurement
EOT	Equivalent oxide thickness
HRTEM	High resolution transmission electron microscopy
IL	Interfacial layer
I-V	Current voltage measurements
MBE	Molecular beam epitaxy
MEIS	Medium energy ion scattering
MOCVD	Metal organic chemical vapour deposition
MOSCAP	Metal oxide semiconductor capacitor
MOSFET	Metal oxide semiconductor field effect transistor
NBTI	Negative bias temperature instability
RTA	Rapid thermal anneal
RTO	Rapid thermal oxide
RTP	Rapid thermal process(or)
SE	Spectroscopic ellipsometry
SILC	Stress induced leakage current
TEM	Transmission electron spectroscopy

VASE	Variable angle spectroscopic ellipsometry
XRD	X-ray diffraction
XRR	X-ray reflection
XPS	X-ray photoelectron spectroscopy

Symbols

A	Oscillator amplitude
A_g	Device area
A_n	First Cauchy parameter
B	Oscillator broadening
B_n	Second Cauchy parameter
C	Capacitance
c	Speed of light in vacuum
C_o	Capacitance density
C_{acc}	Accumulation capacitance
C_e	Capacitance due to electron charge
C_{FB}	Flat-band capacitance
C_{it}	Capacitance due to interface trapped charge
C_{Max}	Maximum capacitance = accumulation capacitance in min-max method
C_{Min}	Minimum capacitance = inversion capacitance in min-max method
C_n	Third Cauchy parameter
C_{ox}	Capacitance of gate oxide
C_p	Capacitance due to hole charge
C_s	Capacitance of the silicon depletion region
$C_{s,FB}$	Capacitance of the silicon depletion

	region at flat-bands
D	Displacement
d	Spacing between atoms in lattice
E	Photon energy
E_c	Conduction band energy level
E_f	Electric field
E_g	Band-gap energy
E_v	Valence band energy level
h	Planck's constant
I	Intensity
I_D	Drain current
j	Imaginary operator ($j^2 = -1$)
J_g	Gate leakage current density
k_b	Boltzmann's constant
k_{hi-k}	Relative permittivity of high-k layer
k_n	Extinction coefficient
k_{ox}	Oxide relative permittivity
k_{SiO2}	Relative permittivity of silicon dioxide
L	Transistor gate length
L_D	Debye length
M	Molar mass
m_o	Electron rest mass

n	Refractive index
N_A	Dopant density (p-type)
NA	Numerical aperture
N_{AV}	Average electron density
N_{Avag}	Avagadro's constant
N_{BD}	Critical defect density
N_D	Dopant density (n-type)
n_e	Number of electrons per molecule
n_i	Intrinsic doping density
P	Polarisation
P_G	Defect generation rate
q	Electron charge
Q_{BD}	Charge to breakdown
Q_{it}	Interface trapped charge
Q_f	Fixed oxide charge
Q_m	Mobile oxide charge
Q_{ot}	Oxide trapped charge
Q_{OX}	Fixed oxide charge
R	Resistance
R_p	Reflectance parallel to plane of incidence
R_s	Reflectance perpendicular to plane of

	incidence
T	Absolute temperature
T_{BD}	Time to breakdown
t_{DL}	Dense layer thickness
t_{OL}	Over-layer thickness
t_{ox}	Oxide thickness
t_{SiO2}	Silicon dioxide layer thickness
t_{hi-k}	High-k layer thickness
V_{AC}	A.C. bias voltage
V_{DC}	D.C. bias voltage
V_{FB}	Flat-band voltage
V_G	Gate bias voltage with respect to substrate
V_{GS}	Gate to source bias voltage
V_m	Molecular volume
V_T	Device threshold voltage
W	Device channel width
W_m	Metal work function
W_{ms}	Metal silicon work function difference
W_s	Silicon work function
$Y(t)$	Electron wavefunction
α	Absorption coefficient

α_e	Electronic polarizability
α_k	Urbach tail amplitude
α_{max}	Half angle of maximum cone of light which can be emitted or received
β_k	Urbach tail exponent
γ	Urbach band-edge energy
γ_d	Charge distribution in oxide
Δ	Ellipsometry angle (differential phase component)
ε_o	Permittivity of free space
ε_1	Real component of the relative permittivity
ε_2	Imaginary component of the relative permittivity
ε_g	Band-gap energy
ε_r	Relative permittivity
ε_{Si}	Relative permittivity of silicon
ε_{SiO2}	Relative permittivity of silicon dioxide
$\varepsilon_{\alpha\beta}^{0}$	Static permittivity tensor
$\varepsilon_{\alpha\beta}^{\infty}$	High frequency dielectric permittivity tensor
θ_{Bragg}	Bragg angle
θ_e	Critical angle for total external reflection

λ	Wavelength
λ_e	De Broglie wavelength
λ_0	Wavelength in vacuum
μ	Carrier mobility
ρ	Physical density
ρ_{DL}	Dense layer physical density
ρ_{OL}	Over-layer physical density
ρ_q	Charge density
ϕ_f	Fermi potential
χ_e	Electronic susceptibility
Ψ	Ellipsometric angle (Amplitude component)

Chapter 1

Introduction and motivation

"I haven't failed. I have just found 10,000

ways that won't work"

A famous quote by THOMAS EDISON while developing an electric storage battery, which could also aptly describe the field of high-k dielectrics research.

1. Introduction

The aim within this section is to summarise the content of the thesis and list the new techniques that have been developed and key research results that constitute the main contributions of the PhD study.

In chapter one, the subject of scaling is addressed; that is, how the limitations encountered by the requirement for ultra-thin silicon dioxide set an ultimate limit for the 45nm node and beyond. It is also discussed how high-k dielectrics provide an extension to conventional scaling techniques by replacing the standard silicon dioxide gate oxide with materials with a higher relative permittivity.

Chapter two covers the theory, fitting techniques and parameter extraction for non-destructive thin film measurements using spectroscopic ellipsometry (SE) which is used extensively in this work. On similar lines, chapter 3 contains descriptions of other analytical techniques which have been used in the work, namely the electrical characterisation techniques of capacitance-voltage (CV) and current-voltage (IV) on MOS capacitors, together with the physical techniques of transmission electron microscopy (TEM), x-ray diffraction (XRD) and x-ray reflectometry (XRR).

Chapter four opens with a review of hafnium oxides and silicates and then results for hafnium oxide and hafnium silicate films are reported. These materials are the main competitors for the replacement of silicon dioxide as the gate oxide for the 45nm technology node, and are in fact, already integrated into Intel products. The chapter first covers ellipsometry measurements of thin MOCVD hafnium oxide/silicate layers and includes extraction of dielectric stack thicknesses, optical constants and band gaps for all of the layers, and in particular, showing how they change with silicon content in the layer. Results in the infra-red regime, are used to de-convolute contributions to the optical constants and show possible composition control problems during deposition. These results were successfully correlated with those from medium energy ion scattering (MEIS) measurements.

A new technique has been developed, for physical density estimation based on the modified Clausius-Mosotti equation. The relationship between silicon

content in the hi-k layer and the physical density was studied. Results are compared with another physical density measurement technique, namely XRR. Finally electrical measurements were carried out and an optimum composition for a hafnium silicate high-k replacement layer was proposed.

Chapter five begins with a review of oxides from the lanthanide series which serves to establish the motivation for considering gadolinium oxide as the focus for the project. It was found that silication of the gadolinium oxide layer proved beneficial to establishing acceptable values for the key metrics of leakage current and capacitance equivalent thickness (CET), within a manufacturing process. The silication was realised by deliberate consumption of the ultra-thin interfacial layer (IL), which happens naturally during the drain activation anneal step in a CMOS manufacturing process. An etching study was performed on thin thermal silicon dioxide to establish a process to control IL thickness and hence to allow the study into the effect of IL thickness on e-beam evaporated gadolinium oxide film properties. Residual oxygen in the processing chamber was found to be a limiting factor and so an annealing process was developed to avoid this occurrence. The CET and leakage current density of silicated, e-beam evaporated gadolinium oxide are shown to meet the target requirements for the 45nm technology node, set within the 'Academic Cluster' work package 2.3 of the PULLNANO European project. The work within the thesis played a significant role in establishing this technology. The Academic cluster comprises of The University of Liverpool, Departments of Electrical Engineering & Electronics together with the Materials Science and Engineering Department, The Tyndall National Institute (Eire), AMO GmbH (Germany) and Chalmers Institute of Technology (Sweden) who collaborate under the name "High-k gang".

Atomic layer deposition (ALD) is seen as a more industrially viable deposition method so two methods of introducing silicon into the gadolinium oxide layer were developed, in line with the experiments on e-beam evaporated samples. A study was undertaken to optimise the process using spectroscopic ellipsometry analysis as a measurement tool. The growth rates and optical constants were studied and the relationship between oxide band gap and deposition parameters

was established. The density relationships were also extracted using the physical density estimation method and correlated to the silicon content.

The samples were also analysed using MEIS, TEM, CV/IV and XRD and conclusions drawn from considering the relationships between electrical properties and annealing temperatures and silicon content. CET values for the ALD deposited stacks were found to be inferior to those of the e-beam samples. It is suggested that this disparity relates to problems with precursor growth instabilities which resulted in thicker than expected high-k layers and also due to the presence of residual oxygen during annealing.

Chapter 6 concludes the study and contains a discussion and summary of the work and results detailed in the thesis. Finally, future work is proposed.

1.1 Scaling

The success of complementary metal oxide semiconductor (CMOS) integrated circuit technology has much to do with the ability of the silicon based microelectronics industry to solve many technological requirements and so allow expansion into more markets than originally thought such as telecommunications, mobile computing, photography and video game entertainment. These requirements include increased speed, low off-state power and the ability to operate with a large range of power supply and output voltages. The method by which this has been achieved was via a process called 'scaling' [1-3]. Scaling is a term which describes the calculated reduction in the dimensions of the elemental device upon which CMOS technology is built, namely the MOSFET. Scaling has allowed CMOS to expand into both the markets of high power, high performance communications and also low power applications. It is true to say that it is the extremely good properties of the gate dielectric, SiO_2 which has been used for CMOS technology since its inception, which has allowed the technology to scale to the dimensions seen today. If a list of the requirements of the ideal gate dielectric were proposed they would be that the film:

- Could be grown repeatedly and uniformly
- Would contain none or minimal oxide charge and defects
- Would be a good diffusion barrier and be thermally stable

- No leakage current under operating conditions
- Large breakdown voltage and time to breakdown

If the properties of SiO_2 are then considered with these in mind, then it is easy to conclude that it is a remarkable material being electrically and thermally stable, with a high quality Si-SiO_2 interface and also has good isolation properties [4]. Furthermore SiO_2 grown within a modern CMOS process, can routinely be produced with low defect charge densities of $\sim 10^{10} cm^{-2}$, mid-gap interface state densities of $\sim 10^{10} cm^{-2} eV^{-1}$ and breakdown fields as high as 15MV/cm [4]. Achieving comparable properties from other replacement oxides presents a substantial challenge for the semiconductor industry and academia.

The need for scaling is a result of the continual market demand for increased functionality and higher performance at a cheaper cost. This has resulted in the semiconductor industry continually attempting to increase the density of chips manufactured on a single wafer [2]. The International Technology Roadmap for Semiconductors (ITRS) is a service sponsored by the leading companies in the semiconductor industry to predict the scaling requirements for future CMOS generations. Every two years the panel of experts involved in this service publish the ITRS report predicting the scaling requirements for the semiconductor industry for up to 15 years into the future. These scaling requirements relating to the gate dielectric for low standby logic as reported in the ITRS 2008 update are shown in table 1.1 [7].

Year in production	2008	2009	2010	2011	2012	2013	2014	2015
Application specific integrated circuit (ASIC) Metal 1 ½ Pitch (nm) (contacted)	59	52	45	40	36	32	28	25
Lg: Physical Lgate for High Performance Logic (nm)	29	27	24	22	20	18	17	15
Lg: Physical gate length for LSTP								
Extended planar bulk and double gate (DG) (nm)	38	32	29	27	22	18	17	15
Ultra-thin body fully depleted (UTB FD) (nm)						20	18	17
EOT: Equivalent Oxide Thickness								
Extended planar bulk (nm)	1.6	1.5	1.4	1.3	1.2	1.1		
UTB FD (nm)						1.2	1.1	1.0
DG (nm)								1.1
$J_{g,limit}$: Maximum gate leakage current density								
Extended planar bulk (mA/cm^2)	81	94	110	120	140	150		
UTB FD (A/cm^2)						150	170	180
DG (A/cm^2)								190

Table 1.1 - Table showing the ITRS scaling roadmap requirements until the 25nm process node [7]

A major problem with the process of scaling is that the gate oxide must be reduced in thickness and the CMOS process is now reaching gate oxide dimensions which both threaten device lifetime and result in unacceptably high values of tunnelling related, gate leakage current [5]. The semiconductor industry is currently looking into many possible solutions for improving device performance further by using alternate routes to traditional scaling such as high mobility channel materials. The proposed solution to the challenge of scaling the gate oxide is to employ metal gates and high-k dielectrics as a replacement for SiO_2 [6].

1.2 Silicon dioxide

There are two key material properties of SiO_2 which are important for MOS device operation and these are the band-gap together with associated oxide-semiconductor band offsets and its reliability under pro-longed circuit operation.

1.2.1 Silicon dioxide band-gap

It has been reported by Muller et al [9] and Tang et al [10] that the band-gap of SiO_2 is established once two or three mono-layers of SiO_2 are formed. Neaton et al [11] proposed that this could due to the fact that in the first monolayer, the oxygen atoms do not have the full complement of six oxygen neighbours and so cannot form the bulk SiO_2 bandgap. The second monolayer is the first layer to have oxygen atoms surrounded by six neighbouring oxygen atoms. Note that the top interface of the film will also not have six neighbouring atoms due to the termination of the film and so it is expected that there is a minimum of 3 required monolayers of SiO_2 before the film has a full bulk bandgap value. Considering the normal bond lengths of the SiO_2 structure, this requirement translates into a fundamental minimum oxide thickness of 0.7-0.8nm [11]. However above this value it has been reported that there is very little change in the key physical characteristics of SiO_2 such as the conduction band offset between SiO_2 and Si [12] and the effective tunnelling mass [13]. Even though there is not such a change in physical characteristics of SiO_2 below 2nm, the magnitude of the leakage currents are very large. Taur et al [14] reported leakage current densities for silicon dioxide films of thicknesses between 1.0-3.5nm and these can be seen in figure 1.1 below. It is apparent from the figure that a 1.5nm silicon dioxide film cannot meet the ITRS leakage current requirements for the 45nm node using a viable value of operating voltage.

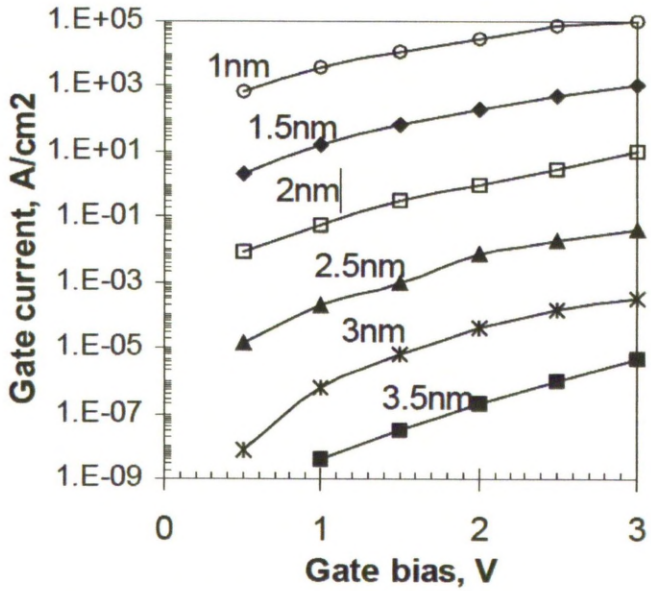

Figure 1.1 - Leakage current densities for ultra-thin silicon dioxide films reported by Taur et al [14]

1.2.2 Silicon dioxide device lifetime

CMOS devices are required to be designed for a working lifetime of at least ten years and so if the dielectric cannot survive this stringent requirement then the technology will be a commercial failure. The breakdown of ultrathin SiO_2 films is thought to follow a percolation model whereby there is creation and build-up of defects within the oxide. After a certain level of electrical and thermal stress arising from normal operating conditions, a complete path of defects can form across the thickness of the oxide [15, 16]. This point defines the breakdown and hence the failure of the oxide, however, self-heating due to high current density within a given percolation path can 'self-heal', due to localised heating and associated annealing-out of the defects [6].

The lifetime of a device is usually quantified in either the charge to breakdown, Q_{BD}, or the time to breakdown, T_{BD}. These are related by the equations below [18].

$$Q_{BD} = {}^{N_{BD}}/_{P_G} \tag{1.1}$$

$$T_{BD} = {}^{Q_{BD}}/_{J_G} \tag{1.2}$$

Here N_{BD} is the critical defect density, P_G is the defect generation under stress and J_G is the gate leakage current density. The critical defect density was extracted by Degrave et al [17]; the defect generation was studied by Stathis et al [18] and it was determined that the charge to breakdown was independent of the oxide thickness for silicon dioxide films of less than 3nm. However the leakage currents for ultra-thin silicon dioxide films increase exponentially in spite of the scaling of supply voltage and this reduces time to breakdown significantly. Using the published values from Degrave [17] and Stathis [18] the time to breakdown has been calculated for silicon dioxide films less than 3nm and the results are shown in figure 1.2 plotted versus the supply voltage.

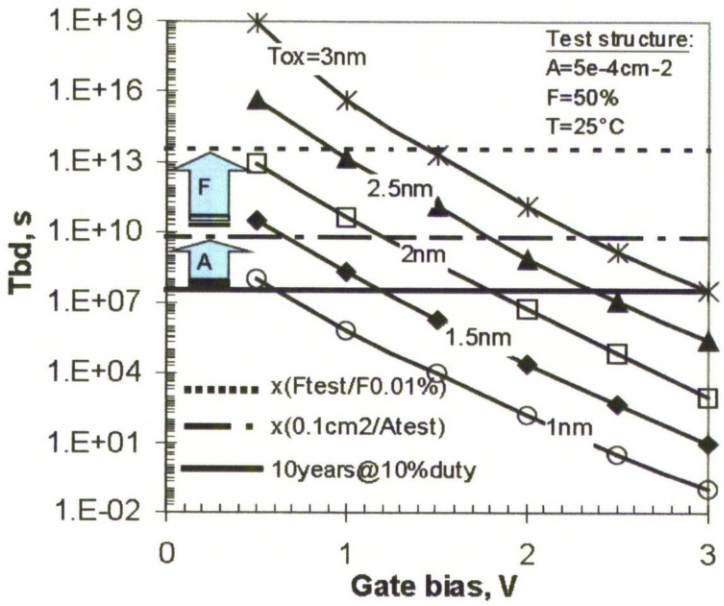

Figure 1.2 - Graph showing the estimated time to breakdown for devices with a silicon dioxide gate dielectric less than 3nm thick plotted versus supply voltage [5]

Skotnicki [5] rescaled the calculations to convert the time to device breakdown, to the time to breakdown for a circuit and predicted that a silicon oxide gate dielectric of 1.5nm cannot meet the requirements for any supply voltage. A thicker film of 2nm was required to allow supply voltages of less than 0.5V, to meet the same 10 year requirements. This condition is not viable due to unavoidable threshold voltage variance. This means that silicon dioxide as a gate insulator cannot meet the time to breakdown requirements for the 45nm technology node which requires an EOT of 1.4nm. The predictions made by

Skotnicki [5] based on the published information by Degrave [17] and Stathis [18] were eventually found to be pessimistic, however, the power dissipation of devices caused by the leakage currents through the thin oxides (see fig. 1.1) was found to be the deciding factor to prevent the use of SiO_2 below the 45nm node.

1.3 Boron penetration

Boron is used to dope the polysilicon material used as a gate electrode in the pMOST transistor in the CMOS process. As the oxide thickness is scaled, significant amounts of boron diffuse through the oxide to the channel during thermal annealing. For gate dielectrics with thicknesses of a few nanometres, diffused boron occurs in unacceptable concentrations within the channel region of MOS devices and hence compromises threshold voltage control [19]. The boron also causes an increase of oxide leakage current. Boron diffusion can be reduced in SiO_2 by the inclusion of nitrogen within the gate dielectric. This inclusion can also bring about other useful attributes such as an increase in the k-value of the dielectric and also increase in the time to breakdown of a device, by reducing hot electron effects and defect generation in the oxide bulk [20-22]. Furthermore, small amounts of nitrogen at the channel interface (~0.1at. %) with the oxide have been shown to improve device performance [23]. Inclusion of nitrogen into the dielectric does however have drawbacks. Any significant amount near the channel interface has been shown to degrade the channel mobility and hence the drive current. This mobility reduction has been attributed to excess charge arising from penta-valent nitrogen atoms, high defect densities due to bonding constraints at the interface (causing increased channel carrier scattering) and from defect levels in the Si-nitride layer which reside near the valence band of silicon [24]. The major reliability issue, negative bias temperature instability (NBTI) can also be significantly compromised if too much nitrogen is present in the border region close to the channel, with in the gate oxide.

Silicon oxynitride dielectrics were seen as the short term solution to the CMOS scaling problems with possible maximum scaling limits to an equivalent oxide thickness of ~1.3nm [25]. Below this value, high gate leakage and reduced

electron channel mobility due to the need for higher channel doping, limit further scaling.

1.4 Other limiting factors for scaling

The physical thickness of the gate oxide differs from the so-called electrical thickness in modern CMOS processes. This is because the electrical thickness is the distance between the centroid of the charge in the channel and the gate electrode. Depletion at the polysilicon gate/oxide interface can occur under normal operating conditions [26]. The depletion region is typically of the order ~0.4nm. The technological challenge is to ensure that the dopant within the polysilicon gate reaches right up to the gate oxide interface, but this in turn establishes a very steep diffusion profile which exacerbates the effect of boron diffusion. Boron penetration is one of the main motivations for the move to metal gate materials as these have a Fermi "sea of electrons" right up to the dielectric interface and so do not experience problems of gate depletion. The other main factor that contributes to the electrical gate thickness for advanced devices is that the presence of high channel doping and ultra-thin oxides results in a quantum mechanical effect whereby the centroid of the channel charge is moved further from the oxide/channel interface. The centroid is positioned at 0.3-0.6nm below this interface within the silicon and effectively, contributes an additional thickness of intrinsic silicon to the dielectric electrical thickness [6].

A further problem relates to the reluctance to reduce supply voltage commensurately with scaling rules for transistor dimensions. The higher supply voltage leads to an increased electric field in the gate dielectric as oxide thickness is scaled. As a consequence, the vertical field in the channel is also increased and so exacerbates surface scattering which reduces channel mobility in the MOSFET on-condition. Physically, the effect arises because increased surface field serves to pull the carriers in the channel closer to the channel/dielectric interface where interfacial roughness causes carrier scattering. The higher operating voltage also has a major impact on power consumption which represents the major issue for further increase in clock speeds and integration, in digital CMOS [6].

1.5 Alternate solutions to scaling

The simple equation for the drive current of a long channel MOSFET for the operating condition after pinch-off ($V_{DS} \geq V_{GS}-V_T$), when the drain current, is saturated is shown below [4]

$$I_D = \frac{W}{L}\mu C_0 \frac{(V_{GS}-V_T)^2}{2} \tag{1.3}$$

The equation describes the on-current, I_{ON} of the MOSFET and is proportional to the speed of a circuit through the relationship, CV_{DD}/I_{ON} where C is the capacitance that must charged and discharged during the switching process and V_{DD} is the supply voltage. This equation can be studied for parameters which could be engineered to set the drain current of the device. The first that could be considered is the factor ($V_{GS}-V_T$). This factor is limited by reliability constraints, namely time to breakdown, which limit the magnitude of V_{GS} (= V_{DD}) and due to the inability to reduce V_T below about 200mV. The latter constraint arises from variations in typical operation temperatures (0-100°C), causing statistical fluctuations in thermal energy, which could adversely affect the value of V_T [6]. The value for mobility, μ is assumed constant for the purposes of this simple situation and so the alterations which would result in an increase of drain current is an increase in capacitance or an increase in the ratio between the gate width and length, (W/L).

Traditionally scaling meant reducing of the length of the channel, but this in turn, results in drain induced barrier lowering (DIBL) and so to avoid this effect the substrate doping must be increased. This increase in substrate doping in turn increases V_T which is then compensated for by decreasing gate oxide thickness. However, further reductions in oxide thickness cannot continue for many more generations due to the associated exponential rise in tunnelling current, as discussed above. This means that to continue scaling, the only path left is to alter the capacitance.

Considering the vertical dimension, and ignoring quantum mechanical and depletion effects, the MOSFET acts essentially like a simple parallel plate capacitor with the gate capacitance described by equation (1.4) below.

$$C = \frac{k_{ox}\varepsilon_0 A_g}{t_{ox}} \tag{1.4}$$

where k_{ox} is the relative permittivity, or dielectric constant which has a value 3.9 for SiO_2, ε_0 is the permittivity of free space, A_g is the area of the gate and t_{ox} is the oxide thickness. If the dielectric is changed from a layer of SiO_2 of relative permittivity k_{SiO2}, to one with a relative permittivity $k_{hi\text{-}k}$ and thickness $t_{hi\text{-}k}$, then the thickness of the new dielectric would have the same capacitance density as for an equivalent thickness of SiO_2. This expression for equivalent oxide thickness (EOT) under these circumstances is given by equation (1.5) below [6].

$$EOT = \frac{k_{SiO2}}{k_{hi-k}} t_{hi-k} \tag{1.5}$$

This means that if the dielectric in a MOS capacitor is changed from SiO_2 (k=3.9) of thickness 2nm to HfO_2 (k=30) then the physical thickness of the HfO_2 can be made to be over 15nm with the same capacitance density. The thicker oxide results in a significantly reduced tunnelling current. Dielectric films thinner than about 4nm, suffer from direct tunnelling, which results in increasingly large values of leakage current density with reducing thickness, as was seen in figure 1.1. Also this equation means that if the physical thickness of replacement HfO_2 is the same as the SiO_2 then the replacement HfO_2 will have the same electrical thickness as a SiO_2 layer scaled to 0.26nm but with improved insulation properties.

1.6. Conclusions

The contents of the thesis and main contributions have been summarised in this chapter. The motivation for the work has been described through a discussion of scaling. The case for replacing the gate oxide material from the native oxide of Si to a thicker, high-permittivity one has been made.

1.7. References

[1] R. H. Dennard, F. H. Gaensslen, H.-N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, IEEE J. Solid-State Circuits, vol. SC-9, pp. 256-268, (1974).

[2] G. Baccarani, M. R. Wordeman, and R. H. Dennard, IEEE Trans. Electron Devices, vol. 31, pp. 452-462 (1984)

[3] P. A. Packan, Science, vol. 285, pp. 2079-2080 (1999).

[4] T. Hori, *Gate Dielectrics and MOS ULSIs,* Springer, New York, (1997)

[5] T. Skotnicki, Proceedings of ESSDERC 2000, pp. 19-33 (2000)

[6] G.D. Wilk, R.M. Wallace and J.M. Anthony, J. Appl. Phys., vol. 89, No. 10, pp. 5243-5275 (2001)

[7] The International Technology Roadmap for Semiconductors, http://public.itrs.net (2008 update) (Web address correct at Sept. 2009)

[8] R. Rios and N. D. Arora, IEDM Technical Digest, pp. 613-616 (1994)

[9] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt and G. Timp, Nature, vol. 399, pp. 758-761 (1999).

[10] S. Tang, R. M. Wallace, A. Seabaugh and D. King-Smith, Appl. Surf. Sci., vol. 135, pp. 137-142 (1998)

[11] J. B. Neaton, D. A. Muller and N. W. Ashcroft, Phys. Rev. Lett., vol. 85, pp. 1298-1301 (2000)

[12] B. Brar, G.D. Wilk and A.C. Scabaugh, Appl. Phys. Lett., vol. 69 pp. 2728-2370 (1996)

[13] B. E. Weir, P. J. Silverman, M. A. Alam, A. Hamad, F. D. Baumann, G. Timp, A. Ghetti, Y. Ma, M. Brown, and T. Sorsch, IEDM Technical digest, pp. 437-440, (1999).

[14] Y. Taur & E.J. Nowak, IEDM Technical Digest, pp. 215-218, (1997)

[15] R. Degraeve, G. Groeseneken, R. Bellens, M .Depas, H.E. Maes, IEDM Technical digest, pp. 863-866, (1995)

[16] R. Degrave, L. Larcher, A. Paccagnella, A. Scarpa, G.I. Ghidini, IEEE Trans. Electron Devices, vol. 45, pp. 904-911, (1998)

[17] R. Degraeve, G. Groeseneken, R. Bellens, M. Depas, H.E. Maes, IEDM Technical digest, pp. 863-866, (1995)

[18] J.H. Stathis and D.J. DiMaria, IEDM Technical Digest, pp. 167-170, (1998)

[19] M. Cao, P.V. Voorde, M. Cox and W. Greene, IEEE Electron Device Lett., vol. 19, 291-293, (1998)

[20] S.V. Hattangady, R. Kraft, D.T. Grider, M.A. Douglas, G.A. Brown, P.A. Tiner, J.W. Kuehne, P.E. Nicollian, M.F. Pas, IEDM Technical Digest, 495-498, (1996)

[21] Y. Wu and G. Lucovsky, IEEE Electron Device Lett., vol. 19, pp. 367-369, (1998)

[22] X.W. Wang, Y. Shi and T.P. Ma, VLSI Symp. Technical Digest, pp. 109-110, (1995)

[23] H. Yang and G. Lucovsky, IEDM Technical Digest, pp. 365-372, (1999)

[24] G. Lucovsky, Y. Wu, H. Niimi, App. Phys. Lett., vol. 74, pp. 2005-2007, (1999)

[25] S. Seungheon W.S. Kim, J.S. Lee, T.H. Choe, J.H. Choi, M.S. Kang, U.I. Chung, N.I. Lee, K. Fujihara, H.K. Kang, S.I. Lee, M.Y. Lee, et al., VLSI Symp. Technical Digest, pp. 190-191, (2000)

[26] R. Rios and N.D. Arora, IEDM Technical Digest. Int. Electron Devices Meet, pp. 613-616, (1994)

Chapter 2

Spectroscopic Ellipsometry

2. Spectroscopic ellipsometry

2.1. Introduction

Ellipsometry is a contactless, non-invasive optical technique for measuring changes in the polarisation state of light reflected from a surface. It is used predominantly to determine the thicknesses of thin dielectric films on highly absorbing substrates but it can also be used to determine the optical constants of films or substrates. The technique does not measure these quantities directly but rather measures other optical parameters of a sample from which the required quantities can be derived.

2.2. The optical constants

The optical constants of a material are defined as parameters which characterise how a material will respond to excitation by an electromagnetic field. Consider a point inside a material where there exists an externally applied electric field, $\mathbf{E_f}$. Then the polarisation, **P**, can be defined as the additional internal electric field induced at this point by the external field, which is modulated by the properties of the material. Neglecting any non-linear effects, this polarisation is proportional to the external field and is given by equation 2.1 below [1].

$$\vec{P} = \tilde{\chi}_e \varepsilon_0 \vec{E}_f \tag{2.1}$$

where χ_e is the electric susceptibility of the material and ε_0 is the permittivity of free space. The displacement, **D**, can be shown using Maxwell's equations, to be equal to the sum of the externally applied field and the polarisation field as shown in equation 2.2 below [1].

$$\vec{D} = (1 + \tilde{\chi}_e)\varepsilon_0 \vec{E}_f \tag{2.2}$$

The dielectric function of the material can then be defined as the constant of proportionality between the displacement and the electric field as shown in equation 2.3 below where $\underline{\varepsilon}$ is that shown in equation 2.4 [1].

$$\vec{D} = \underline{\varepsilon}\vec{E}_f \tag{2.3}$$

$$\underline{\varepsilon} = (1 + \tilde{\chi}_e)\varepsilon_0 = \underline{\varepsilon_r}\varepsilon_0 = \varepsilon_1 + j\varepsilon_2 \tag{2.4}$$

where ε_r is the relative permittivity, ε_1 and ε_2 are the real and imaginary parts of the relative permittivity respectively and ε_0 is the permittivity of free space. In this analysis, the assumption is that the material is isotropic so the response to an applied electric field is equal in all directions. If the film is anisotropic, then the scalar permittivity values above will have to be replaced by a tensor, described by a 3x3 matrix.

The complex index of refraction, $\underline{n}$, is an important parameter describing the optical properties of a material. It is related to the dielectric function by equation 2.5 below [1].

$$\underline{\varepsilon} = \underline{n}^2 = (n + jk_n)^2 \tag{2.5}$$

When this notation is used then the wavelength of the film, λ of an electric wave in the material is determined by the equation below where, λ_0 is the wavelength in vacuum and n is the index of refraction [1].

$$\lambda = {}^{\lambda_0}/_{n} \tag{2.6}$$

The distance, d, into the material where the amplitude of the propagating wave has been reduced to 1/e of its original value is given by equation 2.7 below where k_n is the extinction coefficient. This is equal to the inverse of the attenuation coefficient and is the reason why the imaginary part of the index of refraction is termed the extinction coefficient [1]. This relationship is used to derive the absorption coefficient which can be seen in equation 2.32.

$$d = {}^{\lambda}/_{2\pi k_n} \tag{2.7}$$

2.3. The Kramers-Kronig relations

The real and imaginary values of both types of optical constants are not independent quantities but rather are inter-related and described by the Kramers-Kronig relations. These relations are derived from the requirement that material cannot respond to an applied electric field prior to the application of the field and are shown in equations 2.8, 2.9, 2.10 and 2.11 [1].

$$n(E) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{E' k(E')}{E'^2 - E^2} dE' \tag{2.8}$$

$$k_n(E) = -\frac{2}{\pi} P \int_0^\infty \frac{(n-1)E'}{E'^2 - E^2} dE' \tag{2.9}$$

$$\varepsilon_1(E) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{E' \varepsilon_2(E')}{E'^2 - E^2} dE' \tag{2.10}$$

$$\varepsilon_2(E) = -\frac{2}{\pi} P \int_0^\infty \frac{(\varepsilon_1 - 1)E'}{E'^2 - E^2} dE' \tag{2.11}$$

Here, E is the energy of the incident radiation, P is the principle value which occurs when E' = E. It should be noted that equations (2.8) and (2.9) refer to the (n,k) descriptive format of the material optical constants and (2.10) and (2.11) refer to the real and imaginary parts of the dielectric function. If the absorptive part of the optical constants is known, then the real part can be determined using one of the equations above, or vice versa. The main implication of the use of the Kramers-Kronig relations in ellipsometry is that if the optical constants for a film are extracted from experimental data by using a technique which is not Kramers-Kronig consistent, then the extracted optical constants must subsequently be examined using a model which is. The real and imaginary parts of the optical constants used in such a model must conform to the above relations if they are to be considered physically realistic. An example of a fitting method which may not be Kramers-Kronig consistent is a point-by-point fit as this directly inverts the ellipsometry data to extract the optical constant for each wavelength measured independently. This type of fit fails to take account of the relationship between the extracted real and imaginary optical constants and so risks losing the link with the physical picture. Further modelling and, or parameterisation of the optical constants removes this risk.

2.4. Variation of optical constants with frequency

The most important effect underlying spectroscopic ellipsometry is the frequency dependence of permittivity. This section contains a treatment of these effects.

The first aspect to consider concerns absorption. For a material under excitation by EM radiation, the frequency of an incident electric wave is interchangeable with the energy of photons. Photon energy is related to the wavelength of a wave by equation 2.12 below [2].

$$\lambda(\mu m) = \frac{1.2398}{E(eV)} \tag{2.12}$$

Dielectric and semiconductor materials are equivalent when discussing optical properties and their spectra can be split into two separate regions. The first region starts from the photon energy where the material begins to absorb and continues to higher energies. This region is caused by the onset of interband transitions above the band-gap energy, E_g, and usually the material will be absorbing for all energies above that at which absorption begins. Near the onset energy, the extinction coefficient and the dielectric permittivity would be expected to increase monotonically with energy, as illustrated in figure 2.1, but a more complicated dispersion with energy can be seen if there are contributions from many different inter-band transitions. An example of an absorption caused by inter-band transition can be seen in figure 2.1 and occurs when an electron in a bound state in the material absorbs energy from a single photon from the incident radiation and jumps to a higher level in the material. Semiconductor and dielectric materials exhibit an energy gap in their band structure with a Fermi level occurring, usually within this gap.

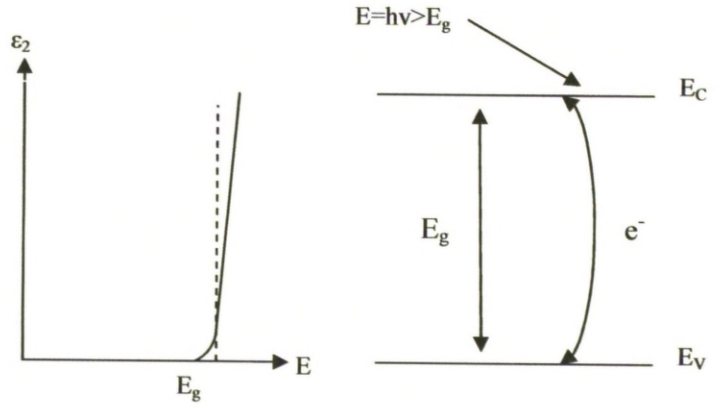

Figure 2.1 - Illustration showing the imaginary permittivity of a simple material and the inter-band transition which occurs to produce this change in permittivity

Therefore, interband absorption cannot occur until the incident photon energy exceeds the energy difference between the highest and lowest occupied energy levels. Below the band-gap energy, the extinction coefficient will be zero for these types of material and the index of refraction would be seen to decrease with decreasing photon energy below the bandap. An important observation is that for regions where the extinction coefficient is zero, there should be no increase of the index of refraction for decreasing photon energy. This property is important when assessing and identifying unphysical results during ellipsometric modelling. In metals, the optical constants of the material are determined by contributions from inter-band absorption as is the case for semiconductor and dielectric materials. However, there are also contributions from intra-band and free electron absorption. As a result, the dispersion of optical constants of metals is usually expected to be complicated.

Intra-band absorption occurs by the process whereby an electron absorbs a photon from the incident radiation but jumps to a different energy state within the same band. An example of an intra-band transition can be seen below.

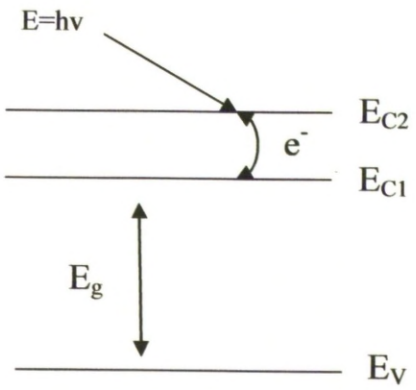

Figure 2.2 - Illustration showing an intra-band transition

The frequency dependence of the absorptive permittivity spectra has been characterised classically by applying a number of relaxation or resonant processes, which all occur around a characteristic frequency. This frequency represents physically, the reciprocal of the characteristic time of the individual

process. This approach is based on the assumption that the material contains electrical dipoles, that is, pairs of opposite charges separated by a small distance. Dielectric relaxation refers to the response of dipoles and electric charges (ionic relaxation) due to an applied alternating field, and can be observed in the frequency range 10^2-10^{10} Hz. Relaxation mechanisms are generally relatively slow compared to resonant electronic transitions or molecular vibrations, which typically have frequencies above 10^{12} Hz. The most common processes are depicted in figure 2.3 and are explained below, starting from the high frequency end of the spectrum.

Electronic polarization refers to a resonant process which occurs in a neutral atom when the electric field displaces the electron clouds relative to the nucleus they surround.

Atomic polarization: A diatomic species of molecule is made up of two atoms, A and B where interaction between them causes a redistribution of electrons along a line AB. This causes a dipole moment along AB unless the atoms are identical in which case, should the dipole be large enough, the material would be described as polar. Under the influence of an applied field, the polarization of a polar substance will change by the atoms being displaced, thus changing the distance between them and hence the dipole moment of the molecule. This is a resonant process, usually occuring in the infra-red region of the frequency spectrum and it is these absorptions which are studied in Infra-Red (IR) ellipsometry.

Dipole relaxation: This is caused by permanent and induced dipoles caused by the same initial conditions described in the process above, undergoing alignment to an electric field. The dipole orientation polarisation is disturbed by thermal noise, which serves to mis-align the dipole vectors from the direction of the field, and the time needed for dipoles to relax which is determined by the local viscosity. These two factors make dipole relaxation heavily dependent on temperature and chemical environment. Hence this is a relaxation process.

Ionic relaxation: This is comprised of ionic conductivity and interfacial and space charge relaxation. Ionic conductivity predominates at low frequencies and

introduces only losses to the system. In a real crystal there are a large number of defects with such examples being lattice vacancies, impurities and dislocations. Free carriers travelling through the crystal under the influence of an applied field can be attracted near to or trapped in these defects and this causes a localised accumulation of charge. This charge causes an image charge on electrodes, thus creating a dipole and this is what constitutes interfacial relaxation.

Referring to figure 2.3, the J.A. Woollam M2000UI spectroscopic ellipsometer used in this work covers the frequency range $1.77x10^{14}$-$1.25x10^{15}$ Hz. The frequency range of current CMOS electronics occurs between 10^{8} – 10^{10} Hz. Interestingly, the two components that contribute most to the optical constants in the CMOS frequency range are the dipole relaxation (due to orientation) and the atomic (or ionic) polarization. Secondly, ellipsometry does not measure the optical constants in the same frequency region as the films would operate in CMOS devices.

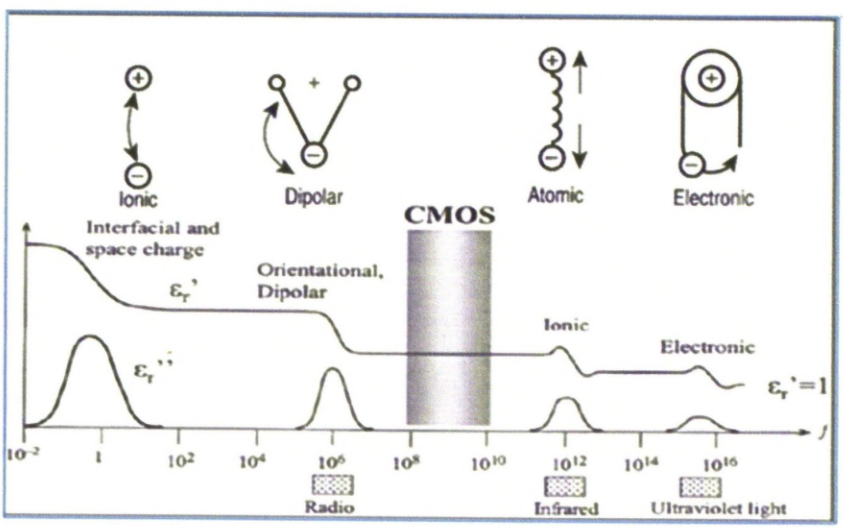

Figure 2.3 - Diagram showing the current CMOS frequency range and which polarization/relaxation processes contribute to the optical constants in the frequency range between 10^{-2} and 10^{16} Hz [4]

2.5. The ellipsometric angles

It is often inconvenient in ellipsometry, to consider separately, how the two different directional components of light reflect at a surface. This is because such an approach does not easily allow the consideration of interactions in multi-layered stacks. Consequently, ellipsometric angles are used to describe how the two reflectance components (p and s) change in relation to each other. The ellipsometric angles are described by the equation below [1]

$$\underline{\rho} = \underline{R}_P / \underline{R}_S = \tan(\Psi)\exp(j\Delta) \tag{2.13}$$

In this equation, R_P and R_S are the complex reflectance coefficients as described by the Fresnel equations in the parallel and vertical components. It can be seen that ψ is an angle which is related to the amplitude of their ratio and that Δ is related to the differential phase shift between the components as further emphasised below [1]:

$$\Psi = \tan^{-1}\left(\underline{R}_P / \underline{R}_S\right) \tag{2.14}$$

$$\Delta = \Delta_P - \Delta_S \tag{2.15}$$

Here Δ_P is the phase shift of the parallel component and Δ_S is the phase shift of the vertical component.

2.6. Incident plane description

Before physical effects are discussed, it is useful to consider the situation of a beam of light incident upon a surface. The light has a 'plane of incidence', which is perpendicular to the sample, as depicted in figure 2.4 which serves to illustrate the orientation. The polarisation state of the incident light is described in relation to this plane using the aforementioned components 'p' and 's'. The 'p' component in this notation, describes the component parallel to the plane of incidence and 's' describes the vertical component (s=senkrecht which is German for vertical) or perpendicular to the plane of incidence.

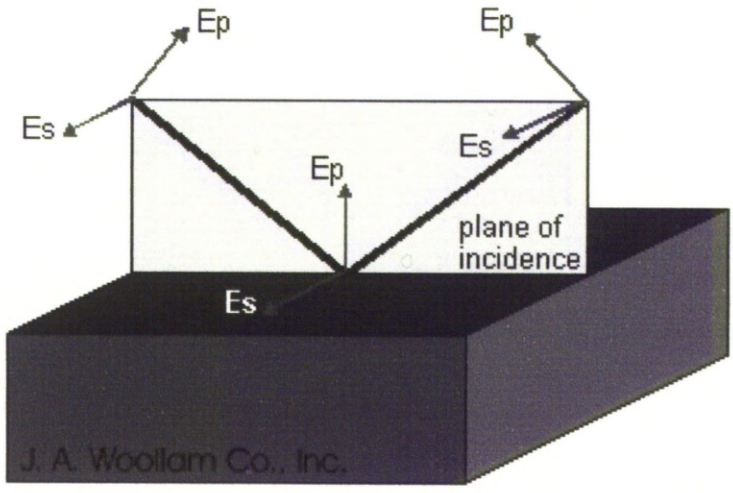

Figure 2.4 - Diagram showing plane of incidence [5]

2.7. Description of incidence angle to reflectance proportionality

It is important in ellipsometry to consider the way in which the two components of the incident beam (p and s) are reflected depending on the angle of incidence. This point is illustrated in figure 2.5, which shows the reflectances for both components for a silicon dioxide film and an aluminium film. Figure 2.5a shows the reflectance of the light as a function of angle of incidence for SiO_2 were 0° corresponds to the angle of normal incidence and 90° corresponds to a beam running parallel to the surface of the material (just grazing). It can be seen that both reflectance components (R_s and R_p) are equal at the two points 0° and 90°. The equality at 0° is because both planes are indistinguishable from each other and that at 90° is due to the incident light not actually being incident on the surface and so is just passing parallel across the surface into the detector, giving a reflection reading of 100%.

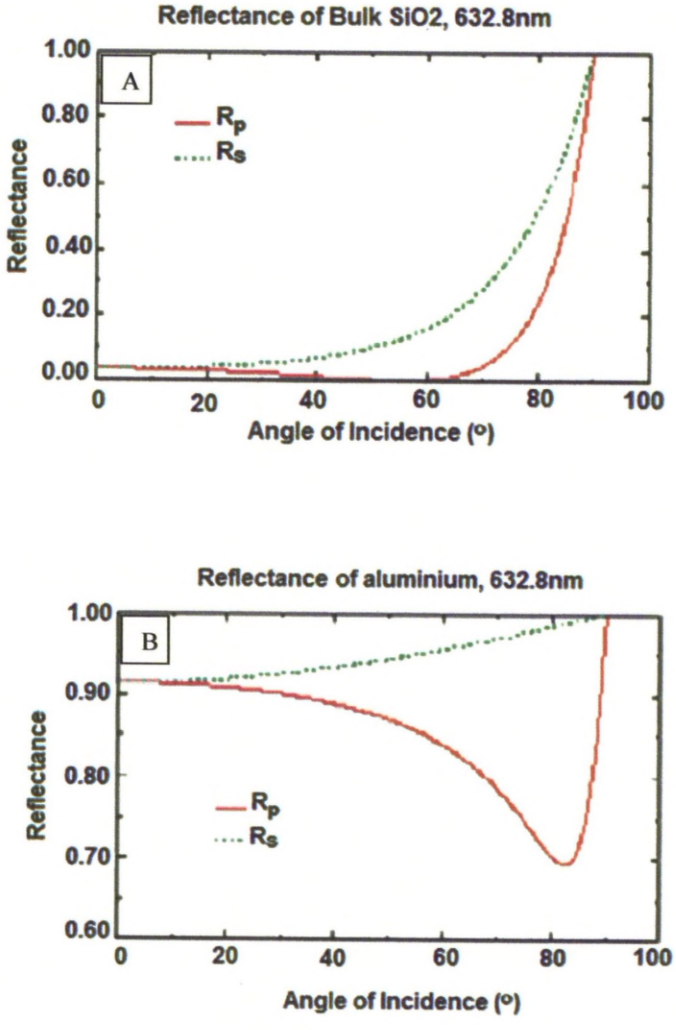

Figure 2.5 - Graphs showing reflectance's for (a) silicon dioxide and (b) aluminium for various incidence angles[1]

Also from the figure, it can be observed that the measured reflectance of both components reacts very differently for both materials, with change of incidence angle. The 's' component increases at an increasing rate towards total incident light reflectance and the 'p' component decreases until it reaches a minimum and then starts to increase rapidly towards the condition of total incident light reflection. The angle at which the p-component is at a minimum is known as the polarisation or the 'Brewster's angle' and it is around the vicinity of this angle that spectroscopic data readings are ideally taken because it is at this angle that the ellipsometry measurements are most sensitive to film characteristics. It

can also be seen from figure 2.5a that the reflectance for both components reaches a significant value for relatively shallow angles. This is because SiO_2 is a quite optically transparent substance.

Turning to figure 2.5b), for an aluminium layer of the same thickness as the SiO_2 one, it can be seen that the plots of the two components have essentially the same basic shape as that of the SiO_2 film. However, the value of reflectance at a normal incidence angle slightly exceeds 90%. This anomaly arises because aluminium is strongly reflective to light at most angles. Another important observation is the fact that the minimum of the 'p' component does not now reach zero. This is due to the high attenuation and reflection coefficient characteristics of aluminium to transmitted light. However, the relevance of the Brewster's angle still stands as this minimum corresponds to that for aluminium of this specification. If the refractive index n_1, of the first medium in which the light was travelling is known, and the refractive index of the medium that is under test is n_2, then the Brewster's angle for a bulk substrate is given by [1]:

$$\phi_B = \tan^{-1}\left(n_2 / n_1\right) \tag{2.17}$$

This relationship holds for a single bulk substrate but most applications of ellipsometry require the analysis of thin films which are layered on a substrate. A different method needs to be used to take account of the Brewster's angle criterion in multi-layered systems, namely the so-called pseudo-Brewster's angle. The pseudo-Brewster's angle takes account of all the layers in the structure. It is initially determined by consideration of the substrate material alone and then modified by taking account of the further layers. It can be shown experimentally that the pseudo-Brewster's angle of a high refractive index substrate with a thin film of a low refractive index, is less than that of the Brewster's angle of the bare substrate. Also it can be shown that a substrate of low refractive index will have a higher pseudo-Brewster's angle if it has a thin film of a high refractive index material.

2.8. Analysis of experimental data

The full ellipsometry measurement results in typically, a large quantity of raw data associated with the intensity and polarisation states of the reflected light over the range of wavelengths selected. These data allow the calculation of the values for ψ and Δ. These values can be used to obtain the film thickness and the relative dielectric constants of the material.

The VASE32 (variable angle spectroscopic ellipsometry) software is used to drive the spectroscopic ellipsometer and analyse the data. This measured data can then be used for physical parameter extraction, such as obtaining film thickness and dielectric functions. A simple model of the sample is assumed with available knowledge of the structure, such as the constituent materials, thicknesses, optical constants and also considerations of sample processing. For the latter, considerations of surface roughness, grading, stoichiometry and porosity are particular concerns. For example if we have measured a sample formed of two layers deposited on a substrate the model would look like that shown in figure 2.6.

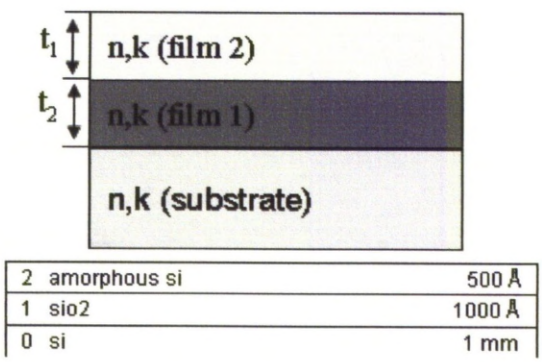

Figure 2.6 – Example of a possible initial model for a sample composed of two layers deposited on a silicon substrate. Note the layers are modelled as layers were thickness and optical functions are the important variables.

The software generates values of psi and delta for the prototype model and this is then compared to that of the experiment. The software incorporates an optimisation routine based on the mean square error (MSE) between the generated and the experimental data. A sufficiently low value for MSE gives confidence as to the accuracy of the model. However, the acceptable MSE

depends on the complexity of the model and hence the measured experimental data. For example, the acceptable MSE for a thin dielectric film on silicon is < 2, but for a very thick film or complex stack then a MSE of <20-30 may be acceptable. Usually the model needs to be modified to achieve the required MSE. The algorithm associated with the above stated methodology is indicated in figure 2.7, which serves to illustrate the stages in the fitting.

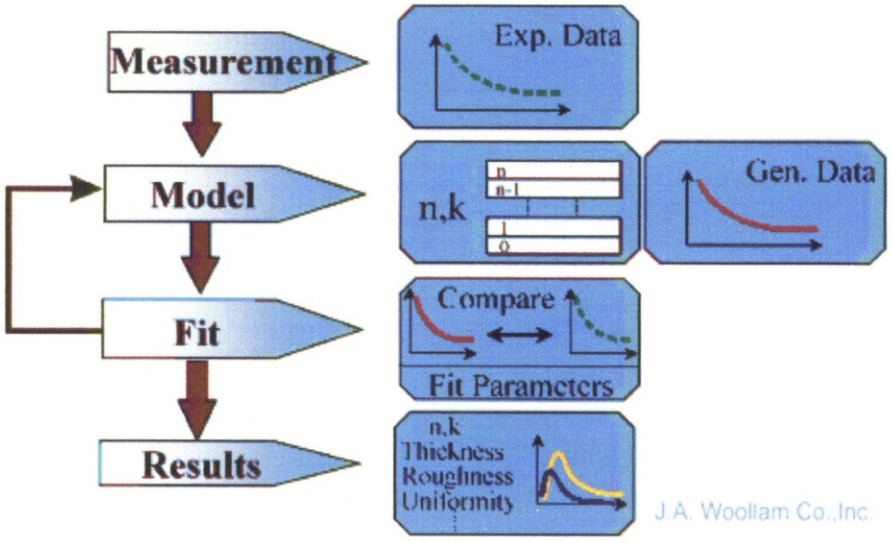

Figure 2.7 – Diagram showing the process of experimental data modelling and important parameter extraction [5]

The choice of initial model is crucial. Figure 2.8 below shows the variation of the MSE one of the thickness parameters is varied. The figure serves to illustrate the well-known problem in optimisation, of false or local minima. The need is to achieve the true or 'global' minimum where the parameter is at its correct physical value. If the thickness corresponding to point A was chosen as an initial value then the algorithm would follow the slope to the global minimum, however if the thicknesses corresponding to points B, C and D was chosen as an initial fit then it can be seen that the probability of the algorithm settling in one of the local minima and giving an erroneous thickness fit value is very high. The problem is compounded by the error value never being exactly zero, due to experimental error and limitations set by the use of relatively simple models applied to represent complex physical situations. As a consequence, it is often difficult to judge when the true global minimum has been found. It is for

this reason, that ultimately, user judgement must be employed in the final analysis. Input from other characterisation methods, particularly TEM, is invaluable both in determining a good initial model, and also reinforcing confidence in the final fit.

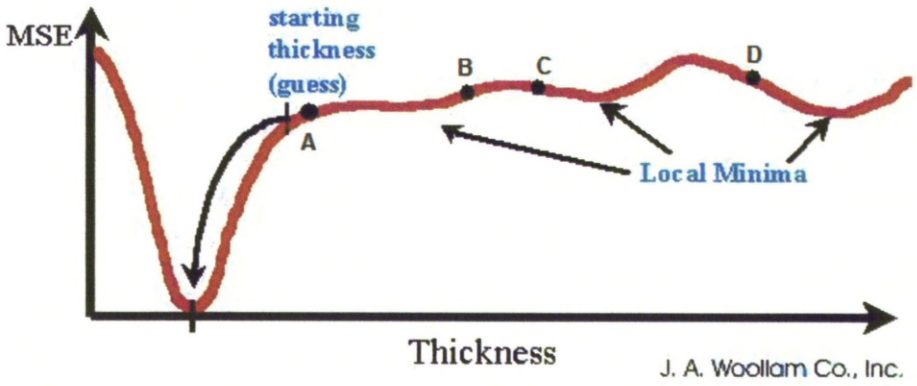

Figure 2.8 - An illustration showing the variations in the mean square error when a particular model parameter is varied [5]

An example of a poor first estimate for a model is shown below in figure 2.9 and an example of the generated and experimental data for a good first estimate model is shown in figure 2.10.

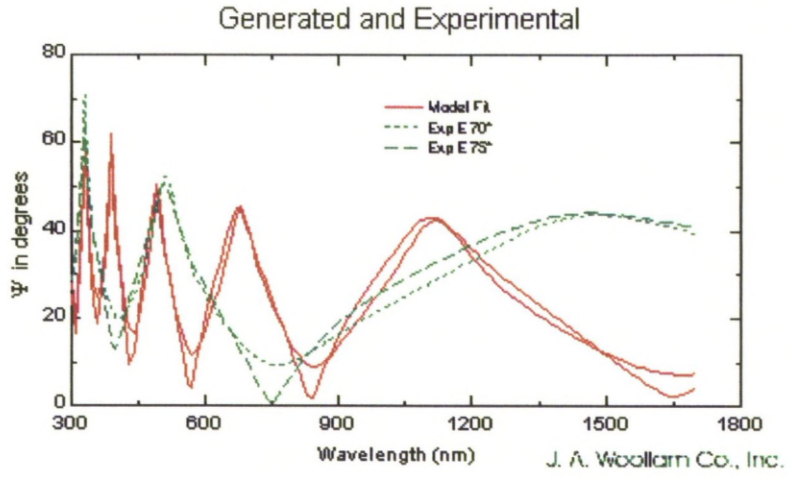

Figure 2.9– Graph showing the experimental values and the generated values of psi for a poor initial model [5]

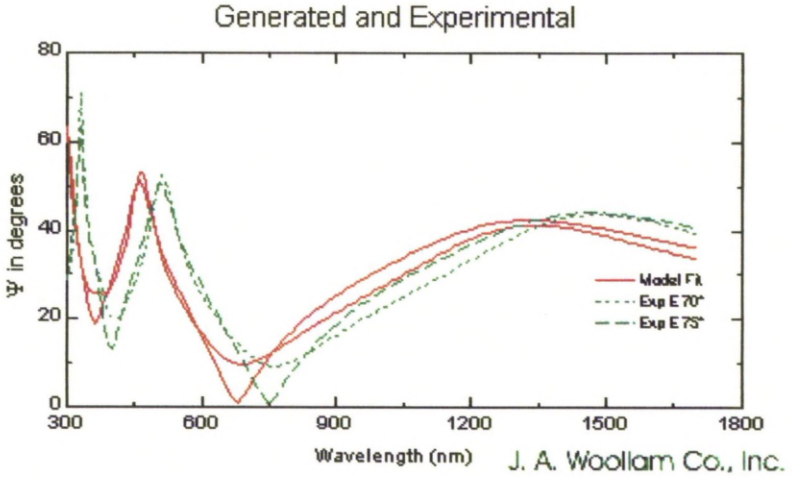

Figure 2.10 – Graph showing the experimental and generated values of psi for a good initial model [5]

2.9. Mean square error (MSE)

As mentioned above, the MSE is a figure of merit, which is used to quantify the difference between the experimental data values and the values generated by the model. The MSE is given by the following formula [1]:

$$MSE = \sqrt{\frac{1}{2N-M}\sum_{i=1}^{N}\left[\left(\frac{\Psi_i^{\mathrm{mod}} - \Psi_i^{\exp}}{\sigma_{\Psi i}^{\exp}}\right)^2 + \left(\frac{\Delta_i^{\mathrm{mod}} - \Delta_i^{\exp}}{\sigma_{\Delta i}^{\exp}}\right)^2\right]} \quad (2.18)$$

where N is the number of (ψ, Δ) pairs in the model, M is the number of variable parameters in the model and σ is the standard deviation of the experimental data points. The 'mod' superscripts indicate data generated from the model and the 'exp' superscripts indicate experimental data.

Expanding on the point mentioned in the previous section, the MSE should not be used as the only quantity for verifying a best fit. The first procedure normally would be to assess by eye, the closeness of the fit to the experimental data to make sure all structure and features are being represented by the model. The optical functions should be examined to ensure they are physically realistic. An example of this would be that the refractive index should be seen to be decreasing monotonically as photon energy decreases, for the case where there is no absorption. The second step would be to examine both the error bars for the fit parameters and the correlation matrix. The error bars for the WVASE program are equal to the figure of merit which is the product of the standard 90% confidence limit and square root of the MSE and this ideally would be considerably smaller than that of the parameter in question (<10%). The correlation matrix describes the inter-relation of the fit parameters; for an accurate model fit, all of the coefficients must typically be below 0.92 [8]. This is because if two parameters are highly correlated then it can be impossible to distinguish independently the correct values for both as altering either would change the generated psi and delta traces in the same way.

2.10. Cauchy layer modelling

The Cauchy layer is the simplest model for representing a layer in the WVASE software. This layer relies upon an inverse power law to model the dispersion in the refractive index for regions where there is no absorption. The dispersion relation for the Cauchy layer is shown below in equation (2.19) where A_n, B_n and C_n are the Cauchy parameters and λ is the wavelength [5]

$$n(\lambda) = A_n + {}^{B_n}/_{\lambda^2} + {}^{C_n}/_{\lambda^4} \qquad (2.19)$$

Figures 2.11 and 12 below show example plots of refractive index versus photon energy according to equation 2.19. It can be seen that an increase in the first Cauchy parameter, A_n, causes an increase in the minimum value for the refractive index in the NIR region. In comparison increasing the second Cauchy parameter can be seen in figure 2.12 to increase the gradient.

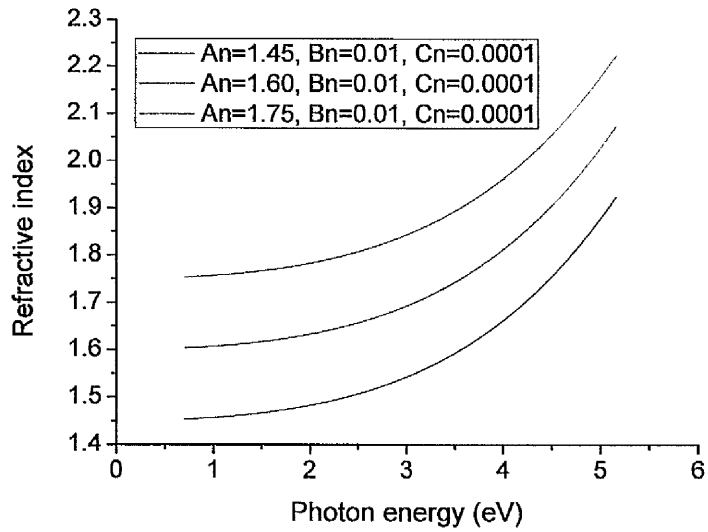

Figure 2.11 - Graph showing the effect of varying the first Cauchy parameter on the resulting refractive index versus the photon energy

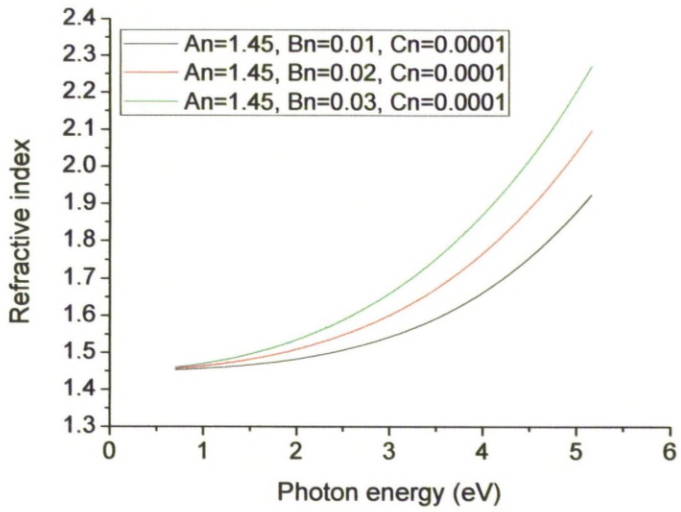

Figure 2.12 - Graph showing the effect of varying the second Cauchy parameter on the resulting refractive index versus the photon energy

If there is a small amount of absorption such as that caused by near band edge defects, then a so-called Urbach absorption tail can be added to the Cauchy model in the WVASE software. The Urbach tail is described by equation 2.20 where α_k is the amplitude of the absorption, β_k is the exponent of the absorption and γ is the band edge of the absorption. In figure 2.13, the extinction coefficient caused by an Urbach tail is shown plotted versus the photon energy for varying values of absorption amplitude. It can be seen that the gradient is increased linearly. Varying the value of the exponent of the Urbach absorption tail can be seen in figure 2.14 to also increase the gradient of the extinction coefficient in an exponential fashion, as expected from equation 2.20. Finally it can be seen in figure 2.15 that changing the band edge value, γ shifts the plot of $\ln(k_n)$ versus photon energy laterally by the change in band gap value [5].

$$k_n(\lambda) = \alpha_k e^{\beta_k\left[12400\left(\frac{1}{\lambda}-\frac{1}{\gamma}\right)\right]} \tag{2.20}$$

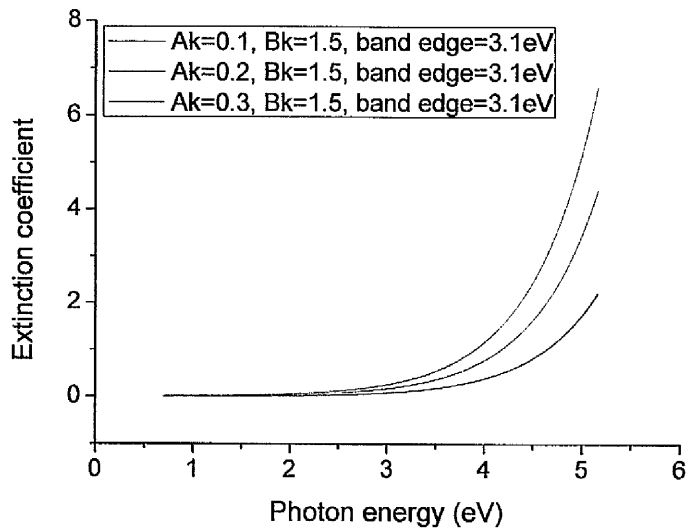

Figure 2.13 - Graph showing the effect on the extinction coefficient versus the photon energy from varying the amplitude of the Urbach tail

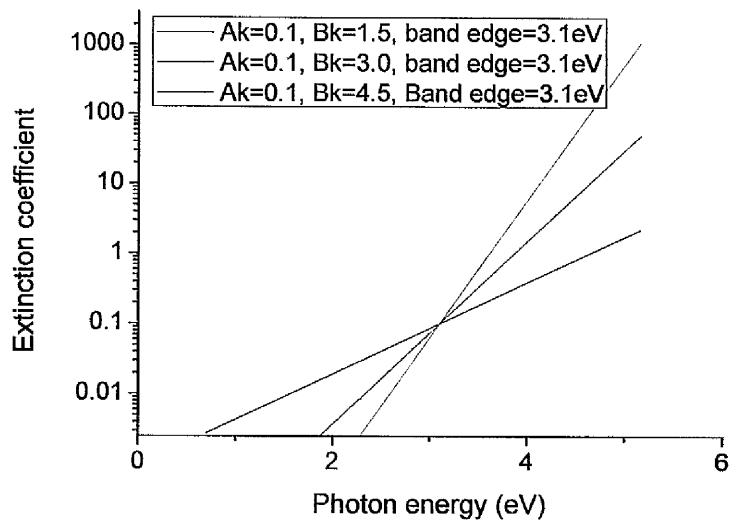

Figure 2.14 - Graph showing the effect of varying the exponent value of the Urbach tail on the extinction coefficient plotted versus the photon energy

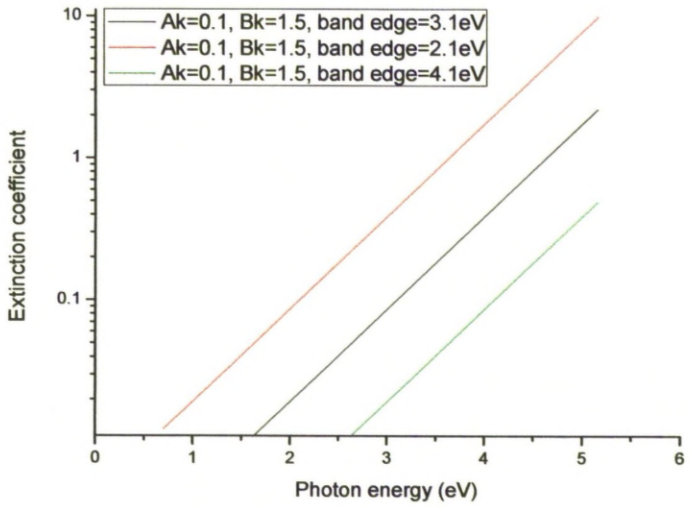

Figure 2.15 - Graph showing the effect of varying the band edge value of the Urbach tail on the extinction coefficient plotted versus the photon energy

The theory of light-matter interaction upon which the Cauchy distribution is based has no physical basis. However, it can be empirically shown to model the dispersive shape exhibited by the refractive index of a material in wavelength regions where there is no absorption. Furthermore, it is only valid for regions of normal dispersion in the visible wavelength region as it cannot account for atomic absorption in the infra-red and electronic absorption in the deep ultraviolet. However, it is particularly useful in ellipsometry data analysis due to its simplicity and proven accuracy over the wavelength ranges measured by UV-Vis-NIR ellipsometry.

The process of fitting a Cauchy layer to ellipsometry data follows the following procedure:

1. The transparent region of wavelength is selected,
2. The first Cauchy parameter (A_n) is adjusted until the amplitude of the generated psi values compare in amplitude to the experimental psi values,
3. The layer thickness parameter is adjusted to match the number of oscillations,

4. The values of thickness and A_n are fitted,
5. The parameter B_n is included in the model and a further fit undertaken,
6. The parameter C_n is included in the model for a final re-fit. If the MSE reduces by > 10% then C_n is left in the model, if not then the parameter is removed.
7. Finally, if Urbach absorption is to be included, then the wavelength range for the fit is expanded to the whole measured range, the Urbach band edge value is set to coincide with the region where absorption is expected to become dominant and the values for the amplitude and exponent of the Urbach tail are fitted with the Cauchy parameters.

If the transparent region wavelength range is unknown then a very long wavelength region is selected (900-1700nm) and the model is fitted multiple times in step 4, including shorter and shorter wavelengths (800, 700, 600, etc) until the MSE significantly increases (>200-400%) compared to that in long wavelengths. This large increase in MSE signifies an absorptive region within or close to the region selected.

2.11. The general oscillator

2.11.1. Overview

In some situations, the absorption may not follow a simple dispersion pattern. In such cases, the thickness would be extracted in a transparent part of the measured spectrum, using a simple Cauchy layer fitting firstly for just the thickness and the first Cauchy parameter and gradually adding more Cauchy parameters to refine the fit. The thickness would then be fixed and a point-by-point optical constant fit would be carried out over the whole measured spectral range. This allows a fit of both the refractive index and the extinction coefficient independently for each wavelength from the psi and delta data. Typically, for films >10nm, a 'noisy' trace of the optical constants can be used as prototype optical constants for general oscillator parameterisation, which is now described.

In the general oscillator model, the dielectric function of a film is represented by a linear summation of real or complex terms; the 'oscillators'. A variety of

oscillator models are available in WVASE32, including Drude, Lorentz, Gaussian, Harmonic, Tauc-Lorentz, Herzinger-Johs, and more. The imaginary permittivity component, $\varepsilon_2(\lambda)$, shape is described using oscillators with centre energy, E_c amplitude, A and broadening, B. As mentioned earlier in this chapter, due to the need to maintain Kramers-Kronig consistency, the shape of the real permittivity component, $\varepsilon_1(\lambda)$, is given by the oscillators which exist in $\varepsilon_2(\lambda)$. The real permittivity is also dependent upon oscillators which occur outside of the measured spectral range and so this is modelled by two Sellmeier oscillators and a permittivity offset, ε_{offset}. An example of a general oscillator equation is shown below where 'Offset' relates to the high frequency permittivity and this is followed by a Sellmeier (pole) term, then a Lorentz term and finally a Gaussian term [1].

$$\varepsilon = \varepsilon_1 + i\varepsilon_2 = \varepsilon_{offset} + \frac{A_S}{E_S^2 - E^2} + \frac{A_L}{E_L^2 - E^2 - iB_L E} + Gaussian(E, A_3, E_3, B_3) \quad (2.21)$$

2.11.2. The Lorentz oscillator

Referring to equation 2.21, the Lorentz oscillator is a classic, damped harmonic oscillator model given by equation (2.22) below [1] where A_L is the amplitude of the absorption, E_L is the characteristic energy which the oscillator centres around, B_L is the half-power broadening for the absorption and E is the photon energy. The typical shape of the oscillator can be seen in figure 2.16 below.

$$\varepsilon = \frac{A_L}{E_L^2 - E^2 - iB_L E} \quad (2.22)$$

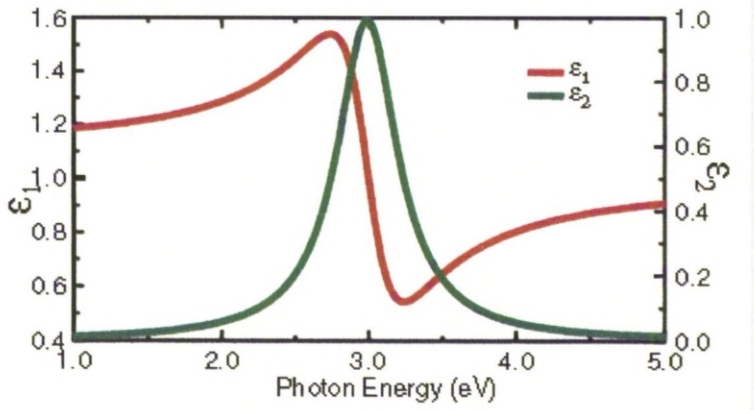

Figure 2.16 - The real and imaginary permittivity's for a Lorentz oscillator with A_L=1, B_L=0.5 and E_L=3eV [5]

The main disadvantage of this type, is that it has long absorption tails which makes it difficult to use when modelling band edge structure in some materials, because the absorption encroaches into the transparent region. It is particularly useful for use in metals however, where absorption exists throughout the measured spectral range.

2.11.3. The Tauc-Lorentz oscillator

The Tauc-Lorentz oscillator was developed by Jellison and Modine [6] and is based on the Lorentz oscillator which was described above. A further problem with the Lorentz model is that the absorption profile is symmetric and this does not occur in amorphous materials as they often show asymmetric shapes. The Tauc-gap of amorphous materials is given below where A_T is the Tauc amplitude and E_g is the Tauc band gap of the material [6].

$$\varepsilon_2(E) = \frac{A_T (E - E_g)^2}{E^2} \qquad \textbf{(2.23)}$$

The Tauc-Lorentz oscillator equation for the imaginary permittivity is derived by multiplying the Lorentz equation by equation 2.23 above and the result is shown in equation 2.24, where A is the amplitude of the oscillator, B is the broadening of the oscillator, E_2 is the characteristic energy of the oscillator, E_g is the bandgap and E is the photon energy [7].

$$\varepsilon_2 = \frac{AEB(E - E_g)^2}{E\left((E^2 - E_2^2)^2 + (B^2E^2)\right)} \qquad (E > E_g)$$
$$\varepsilon_2 = 0 \qquad (E \leq E_g) \tag{2.24}$$

Using the Kramers-Kronig relations, it is then possible to derive an equation for the real component of the relative permittivity and this can be seen in the system of equations overleaf [7].

$$\begin{aligned}\varepsilon_1 = \varepsilon_1(\infty) &+ \frac{AB}{\pi\xi^4}\frac{\alpha_{ln}}{2\alpha E_2} ln\left(\frac{E_2^2 + E_g^2 + \alpha E_g}{E_2^2 + E_g^2 - \alpha E_g}\right) \\ &- \frac{A}{\pi\xi^4}\frac{\alpha_{tan}}{2\alpha E_2}\left[\pi - tan^{-1}\left(\frac{2E_g + \alpha}{B}\right) + tan^{-1}\left(\frac{-2E_g + \alpha}{B}\right)\right] \\ &+ 2\frac{AE_2}{\pi\xi^4\alpha}E_g(E^2 - \gamma^2)\left[\pi + 2tan^{-1}\left(2\frac{\gamma^2 - E_g^2}{\alpha B}\right)\right] \\ &- \frac{AE_2B}{\pi\xi^4}\frac{E^2 + E_g^2}{E} ln\left(\frac{|E - E_g|}{E + E_g}\right) \\ &+ \frac{2AE_2B}{\pi\xi^4}E_g ln\left[\frac{|E - E_g|(E + E_g)}{\sqrt{(E_2^2 + E_g^2)^2 + E_g^2B^2}}\right]\end{aligned} \tag{2.25}$$

Where

$$\alpha_{ln} = (E_g^2 - E_2^2)E^2 + E_g^2B^2 - E_2^2(E_2^2 + 3E_g^2) \tag{2.26}$$

$$\alpha_{tan} = (E^2 - E_2^2)(E_2^2 + E_g^2) + E_g^2B^2 \tag{2.27}$$

$$\xi^4 = (E^2 - \gamma^2)^2 + \frac{\alpha^2B^2}{4} \tag{2.28}$$

$$\alpha = \sqrt{4E_2^2 - C^2} \tag{2.29}$$

The symbols are as defined for the imaginary component above. Although the equation for the real component of the permittivity, ε_1, is quite complex it only contains five parameters. Figure 2.17 below shows a comparison example for the imaginary permittivity of Tauc-Lorentz and identical Lorentz, oscillators, except for the E_g parameter which exists only in the Tauc-Lorentz oscillator. It is apparent that the absorption peak for the Lorentz is symmetric whereas the absorption peak for the Tauc-Lorentz oscillator is tilted by comparison. This tilt increases the gradient of the absorption peak on the lower energy side and

reduces the absorption below the value of E_g to zero, which makes the oscillator well-suited for modelling the band edge of amorphous materials.

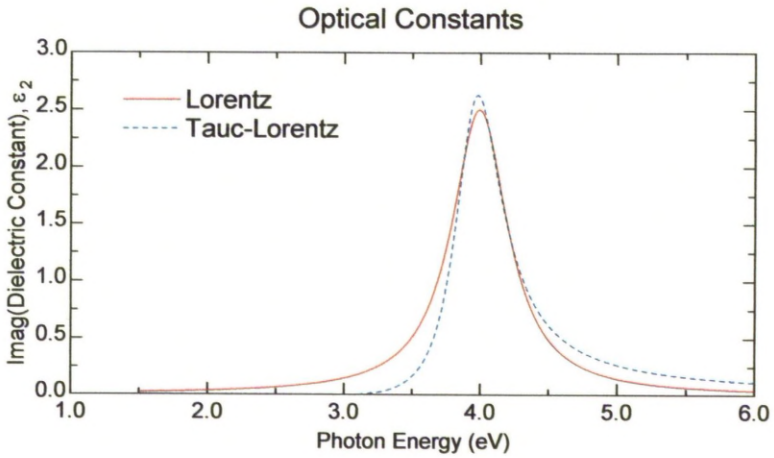

Figure 2.17 - Graph showing the imaginary permittivity for classical Lorentz (A=2.5, B=0.5 and E_2=4eV) and Tauc-Lorentz (A_L=2.5, B_L=0.5, E_L=4eV and E_L=3eV) oscillators

2.11.4. The Sellmeier pole

The Sellmeier (pole) is an optical model based on an empirical relationship between the real component of relative permittivity and the wavelength in a transparent medium and it is given by the equation shown below. In this equation A_S is the amplitude of the pole, E_S is the characteristic energy the pole occurs at and E is the photon energy of the incident light. This produces structure in the real permittivity as shown in figure 2.18 below [5]

$$\varepsilon_1 = \frac{A_S}{E_S^2 - E^2} \tag{2.31}$$

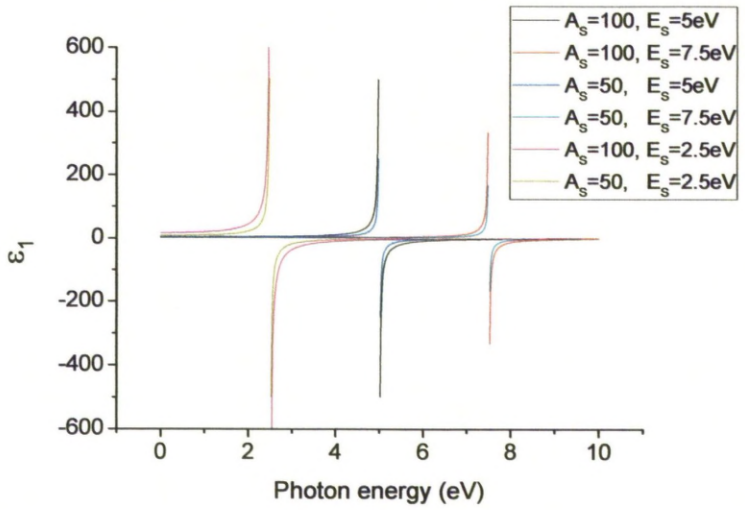

Figure 2.18 - Graph showing Sellmeier poles with various values for amplitude and characteristic energy

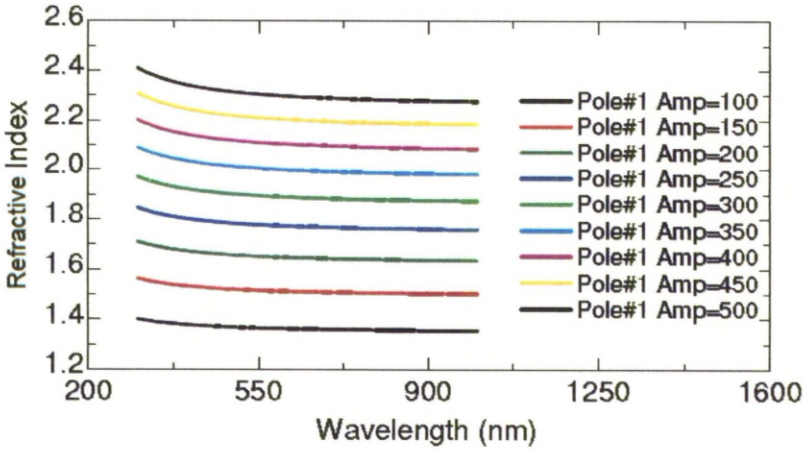

Figure 2.19 - Graph showing the effect of altering the amplitude of a Sellmeier pole with characteristic energy of 11eV on the refractive index

It can be seen from figure 2.18 that the characteristic energy, E_S is the energy at which the inflection point occurs. In the general oscillator layer, two Sellmeier 'poles' are used to model structure associated with absorptions outside the measured spectral range. If the characteristic energy for a Sellmeier pole is fixed and the amplitude is varied, then it can be seen in figure 2.19 that the refractive index trace shifts up in magnitude. In figure 2.20 below it can be seen

that there are absorptions in the low energy spectrum due to lattice vibrations which cause a point of inflection in ε_1. There is also another group of oscillations at higher energies due to electronic absorptions causing another inflection point in ε_1. It is obvious that if two Sellmeier poles are used, then so long as the points of inflection occur outside of the measured range. These can be used to model the residues in the measured spectral range arising from these absorptions. An example is shown in figure 2.21 of two Sellmeier poles being used in this way.

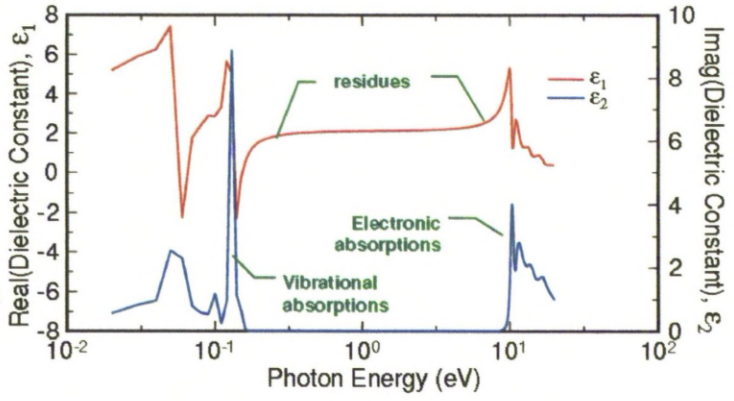

Figure 2.20 - Illustration showing the absorptions outside the measured spectral range [5]

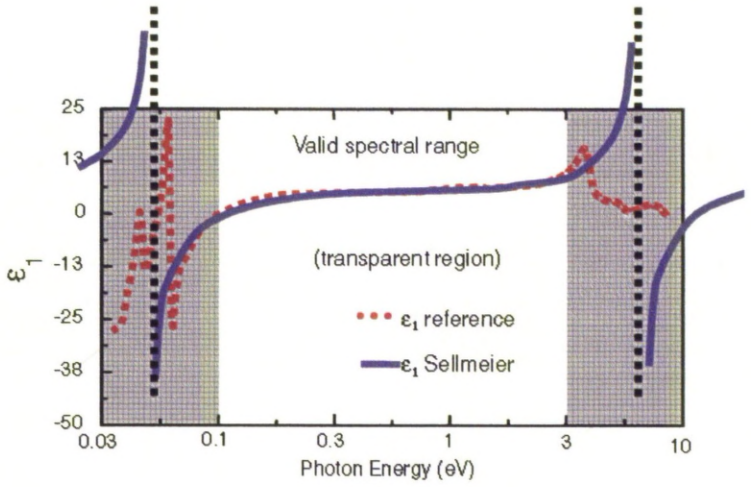

Figure 2.21 - Illustration showing how two Sellmeier poles can model the shape of the relative permittivity in the measured spectral range caused by absorptions outside of the measured spectral range [5]

2.11.5. Fitting process for modelling experimental data with a general oscillator layer

There is a standard procedure for fitting a general oscillator model to experimental data. The first stage is to extract the thickness and prototype optical constants using a Cauchy layer. Stages 1-6 in the Cauchy fit procedure explained earlier are then carried out to extract the correct value of thickness in the transparent region of the measured spectrum. The thickness for the layer is then fixed and a point-by-point optical constant fit is performed for the refractive index and extinction coefficient. This procedure fits the refractive index values and extinction coefficient values for each value of wavelength independently giving 'prototype' optical constants.

The next stage of the fitting procedure is to parameterise the prototype optical constants to remove noise, identify individual contributions and check for Kramers-Kronig consistency. The general oscillator is split into two parts. The first part models absorptions caused by band edges and defects within the measured spectral range by using several oscillator models. The second part is to model the features in the real permittivity due to absorptions far outside the measured spectral range. In figure 2.22 below, a screenshot showing details of the fit using a general oscillator layer can be seen for the parameterisation of an HfO_2 layer with parameters taken from the WVASE database. In this figure, section A relates to the real part of the permittivity, whereas section B details oscillator parameters used to model the absorptions within the measured spectral range. Section C shows the fit of the prototype optical constants to those generated from the model.

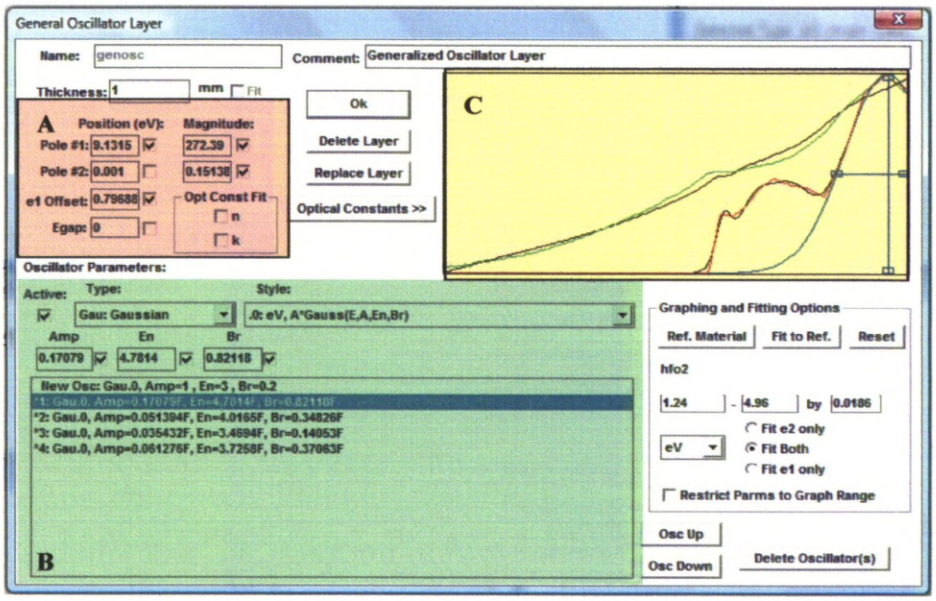

Figure 2.32 - Screenshot of the WVASE32 General oscillator layer showing a) the Sellmeier Model with two poles, b) Multiple oscillator model and c) optical constants fit window

The fitting process is then as follows:

1. The prototype values for the refractive index and the extinction coefficient are loaded into the model and are converted into relative permittivity values.
2. The user then employs as few different oscillators as possible to match the absorption peaks seen in the imaginary component of the permittivity. A fit of the oscillator parameters to the prototype imaginary permittivity is then done to optimise their values. In figure 2.22, four Gaussian oscillators were found to be sufficient to model the absorption peaks, with some confidence.
3. The software then calculates the contribution to the real relative permittivity from the oscillator absorption peaks.
4. The contributions to the real permittivity outside of the measured spectral range are then modelled by the two Sellmeier poles. These are fitted to the prototype real permittivity including the effects of the oscillators fitted earlier using the following process:

a. First the magnitude of the high energy pole is fitted,
b. The energy position of the high energy pole is included in the fit,
c. The magnitude of the low energy pole is included into the fit with the parameters of the high energy pole,
d. Finally an offset accounting for the real permittivity is included as a fit parameter.
e. IMPORTANT: the low energy pole energy position should never be fitted.

5. All of the parameters for both the poles and oscillators should be given one last fit together to optimise both the real and imaginary permittivity fits.
6. This is then the final general oscillator model and a fit is then performed by comparing the psi and delta values from this model to the experimental psi and delta values to optimise the model to account for any anomalies between the actual optical constants of the real layer and the extracted prototype (point-by-point) optical constants. Possible anomalies include artefacts due to noise and Kramers-Kronig inconsistencies.

2.12. Challenges with modern ellipsometry

Ellipsometry as a characterisation technique for thin films has been seen traditionally as a technique used for the estimation of thin films thickness and optical constants. The recent developments in both hardware, particularly variable angles and the extended spectral range, together with data assessment software, introduce new capability to ellipsometry. The characterizations of high-k dielectric films pose a series of challenges most notably:

a) Ultra thin films (<4 nm thickness) with a structure dependent on the growth or deposition technique used.
b) Multi - stack structures, due to the presence of a transition layer at the interface between substrate and the metal oxide film and a possible variation in the layer composition.

c) The lack of reliable information on optical constants for most ultra-thin, novel materials.

The biggest obstacle in correctly analysing a modern CMOS gate dielectric is related to the ultra-thin nature of the gate stack. Ellipsometry measurements of a thick film display interference oscillations which have different amplitudes, positions and broadening depending on the physical and optical properties of the measured film, as shown in figure 2.24. It is known for instance, that for a Cauchy layer on a silicon substrate, the first Cauchy parameter is related to both the position and amplitude of the peaks in psi and that its thickness is only related to the position of the peaks and psi. This means that the optical constants and film thicknesses can be extracted independently and that it can therefore be said that there is no correlation between these fitted parameters.

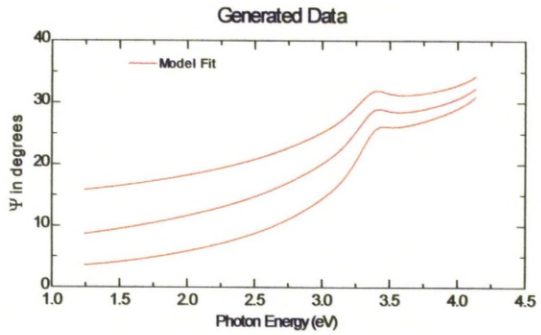

Figure 2.23 - Graph showing generated data for a 10nm silicon dioxide film on silicon

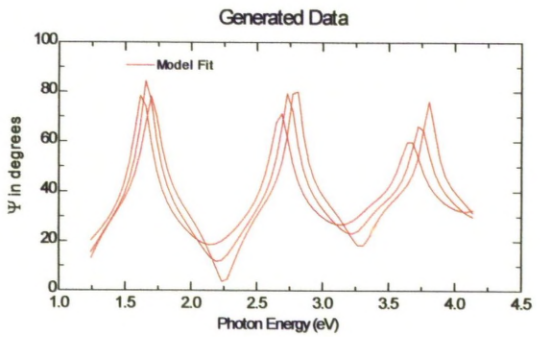

Figure 2.24 - Graph showing generated data for a 500nm silicon dioxide on silicon

The challenge posed by films less than 10nm is that they are too thin to cause interference oscillations and so the psi parameter displays dispersion with minimal features, as shown in figure 2.23. The problem with trying to fit an optical model to a psi trace which displays only a slope is that the gradient and position of this slope can be modelled by many different pairs of thickness and optical constants. In this situation, the thickness and optical constants are said to be highly correlated and it is extremely difficult, if not impossible, to resolve each parameter independently from the ellipsometric data. Considering the situation as film thickness is reduced, the peaks in psi shift to higher energies because the interference for a thinner film affects smaller wavelengths. Then a possible method for reducing the effect of correlation is to perform the ellipsometry using a spectral range covering higher photon energies, where interference oscillations may be present. This is the reason why in chapter three, the very thin hafnium oxide and hafnium silicate films were analysed using the VUV-VASE system, as this measures ellipsometric data for photon energies ~4eV higher in energy than the M2000UI system, hence giving more peak information for the modelling process.

2.13. Band-gap extraction using Tauc-plot

The band gap of a film is an important parameter which can be extracted quite effectively using the optical constant data from ellipsometry. The simplest technique for extracting the band gap of a film uses a so-called Tauc plot. The first step is to calculate the absorption coefficient, α from the extracted extinction coefficient over the measured spectral range using equation 2.32:

$$\alpha(\lambda) = \frac{2\pi k_n(\lambda)}{\lambda} \tag{2.32}$$

The Tauc plot involves plotting the quantity $(\alpha nE)^{1/2}$ versus E. Figure 2.25 shows an example of such a plot and the band gap can be extracted from a straight line drawn parallel to the $(\alpha nE)^{1/2}$ extrapolated to the x-axis axis, at the point of maximum gradient. The photon energy at this intercept point represents an estimate of the band gap for the film. Sometimes two distinct regions with different gradients can be seen and this can be caused by a defect level in the

film. For such a situation, two lines can be drawn to extract both the band gap and the near to band edge, defect level.

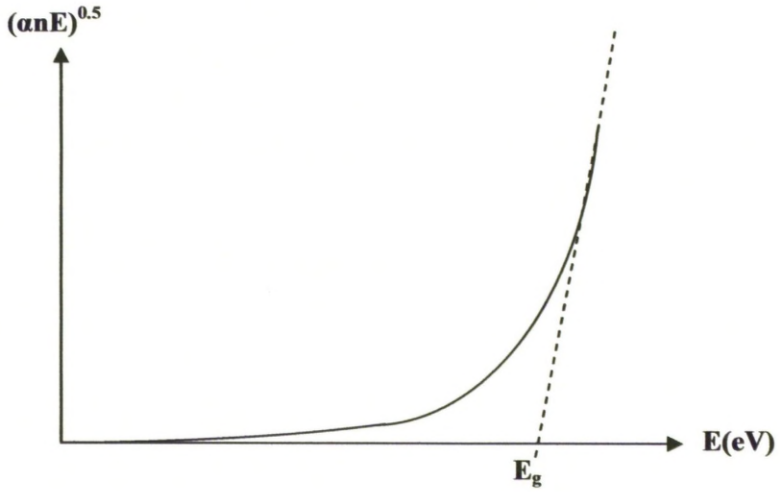

Figure 2.25 - An illustration of a Tauc plot showing an example of bandgap extraction

2.14. Conclusions

In this chapter the optical physics of materials has been introduced and contributions to the refractive indices in different frequency ranges have been explained. The basic principles of spectroscopic ellipsometry were explained and standard fitting procedures were described for both a simple Cauchy layer and a more complex general oscillator layer. It was also shown how the physical properties of a material such as the band-gap and trapping levels could be extracted using the technique. Modern challenges for spectroscopic ellipsometry as an analysis technique for gate dielectrics was also discussed and solutions for some of the problems were proposed.

2.15. References

[1] Instruction manual for the M2000UI spectroscopic ellipsometer, J.A. Woollam®

[2] Semiconductor material and device characterisation, D.K. Schroder, 2nd edition, Wiley, (1998)

[3] Application Note 1217-1, "Basics of measuring the dielectric properties of materials," Hewlett Packard literature number 5091-3300E, (1992).

[4] S.O. Kasap, Principles of Electrical Engineering Materials and Devices, (2nd edition, McGraw-Hill, 2002

[5] Course notes for "Application orientated Woollam WVASE training course". Darmstadt, Gerrmany. Operated by LOT-Oriel and J.A. Woollam (March 2008)

[6] G. E. Jellison, Jr., F. A. Modine, Appl. Phys. Lett. 69, 371-373 (1996) and Erratum, Appl. Phys. Lett. 69, 2137 (1996).

[7] Handbook of Ellipsometry, H.G. Tompkins and E.A. Irene, Springer, (2005)

[8] Private communication with Blaine Johs, J.A. Woollam Inc., Nebraska, USA

Chapter 3

Other analytical techniques

3. Other analytical techniques

3.1. Capacitance-voltage analysis (C-V)

A capacitance-voltage plot is a relatively simple and fast analysis method which can be carried out on a metal oxide semiconductor capacitor (MOSCAP) structure to gain information on many of its electrical properties such as the k-value for the gate oxide, the doping density of the substrate and the flat band voltage shift. The structure of a MOSCAP is as shown in figure 3.1 below.

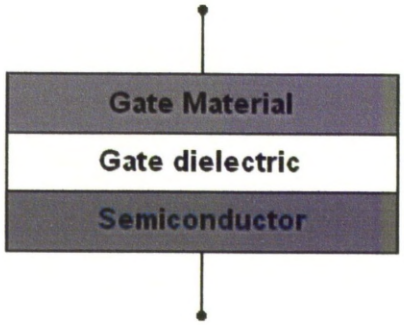

Figure 3.1 - Diagram showing the structure of a metal oxide semiconductor capacitor (MOSCAP)

The capacitance-voltage (C-V) plot can be measured for a MOSCAP by using the simple circuit shown below in figure 3.2. The impedance measuring circuits are much more complex than this in the HP4192A Impedance analyser which was used to measure the C-V plots in this thesis; however, this simple circuit can give an understanding of the measurement procedure.

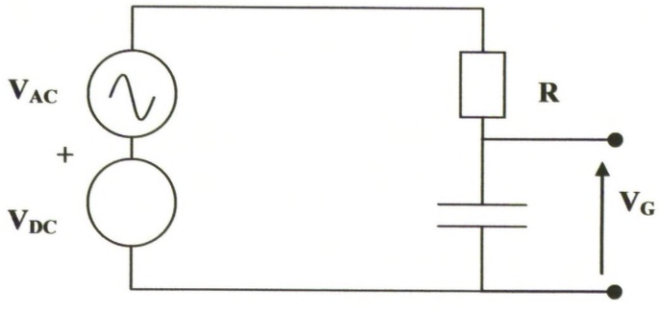

Figure 3.2 – Basic C-V analysis circuitry

In this circuit the AC voltage is kept constant (typically 50mV) and the AC frequency is selected for a particular sweep (typically 0.1 kHz – 1MHz). The DC voltage is then swept slowly across a range sufficient to bias the device from accumulation to inversion with the impedance measured across the capacitor. This impedance can then be converted into a capacitive component and a resistive component using the measured phase angle.

The equivalent capacitive circuit for a MOSCAP device is shown below in figure 3.3 where C_{ox} is the gate oxide capacitance, C_{it} is the capacitance caused by interface trapped charge, C_s is the capacitance caused by the semiconductor depletion region, C_p is the capacitance due to hole charge and C_e is the capacitance due to free charge in the accumulation layer.

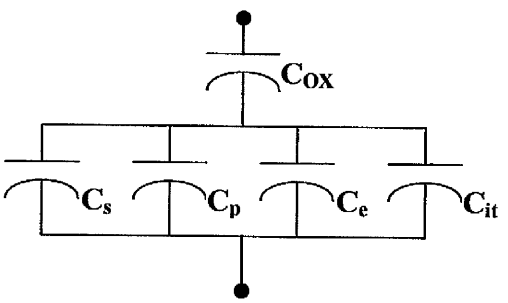

Figure 3.3 - Illustration showing the equivalent capacitive circuit for a MOSCAP device [1]

To analyse a capacitance-voltage plot then it must be first understood the physics which is occurring in the device in all the different bias conditions. If an n-type MOSCAP is considered then with a significant positive bias voltage on the gate, electrons will be attracted to the semiconductor/gate dielectric interface forming an accumulation layer. These majority carriers can follow even high frequencies and so in the equivalent circuit in figure 3.3 the C_e dominates, shorting the other capacitors. Hence the capacitance in accumulation is equal solely to the gate oxide capacitance. If the positive bias is reduced then a point will be reached (Vg=0 in ideal device) where there is no net charge in the device; the flat-band condition as on an energy band diagram the conduction and valence band would be flat. As the gate bias is made slightly

negative then the free carriers will be repelled a distance from the semiconductor/gate dielectric interface creating a depletion region. In this region the free charge is considered to be zero so the equivalent capacitive circuit is the same as figure 3.3 with C_c open circuited.

If the gate bias is made significantly negative then minority carriers (holes in n-type) will be generated by electron hole pair creation in the depletion region, at room temperature. For the case of n-type material, the generated electrons are repelled out of the depletion region by the electric field and the holes drift to the semiconductor/gate dielectric interface, hence inverting the surface. In this region the resulting total capacitance relies on the frequency of the A.C. voltage. If the frequency of the A.C. voltage is low then the minority charge can follow this and so in figure 3.3 C_p will dominate and hence short circuit the parallel capacitors resulting in a total capacitance equal to that of only the gate dielectric resulting in the low frequency C-V curve. If the frequency of the A.C. voltage is too high then the inversion charge will not be able to follow this resulting in the total capacitance being the combination of C_s and C_{ox}, which gives the high frequency C-V curve. The final situation is if the ramp rate for the D.C. voltage is faster than the minority charge can generate and this will result in the deep depletion curve. An example of the high frequency, low frequency and deep depletion C-V characteristics for an n-type sample is shown in figure 3.4.

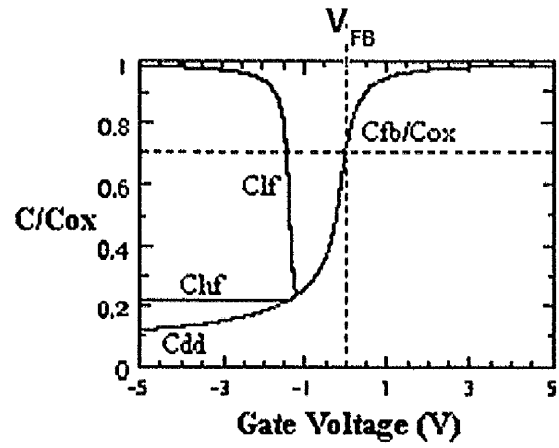

Figure 3.4 - Picture showing a high frequency C-V, Chf, a low frequency C-V, Clf and a deep depletion C-V, Cdd for an ideal n-type sample with $N_D=10^{17}cm^{-3}$, t_{OX}=10nm and T=300K [1]

Above it was discussed what physics occurs within an n-type MOS capacitor for accumulation, depletion, flat bands and inversion. Similar conditions occur in a p-type MOS capacitor, except that the polarities of voltages, charge and band bending are reversed, and the roles of electrons and holes are interchanged (any positive donors discussed above would be interchanged for negative acceptor ions). The C-V sweep for a p-type sample can be seen in figure 3.5 below showing the high frequency, low frequency and deep depletion C-V curves

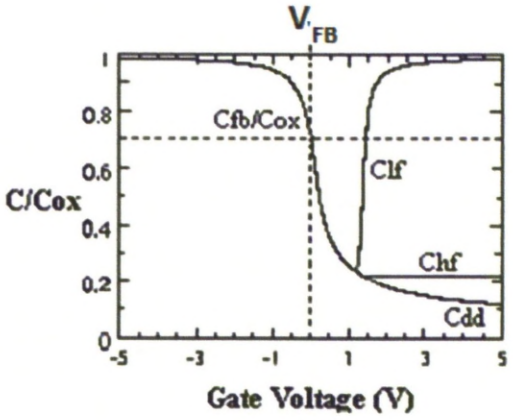

Figure 3.5 - Picture showing a high frequency C-V, Chf, a low frequency C-V, Clf and a deep depletion C-V, Cdd for an ideal p-type sample with $N_A=10^{17}cm^{-3}$, $t_{OX}=10nm$ and T=300K [1]

From above it can be seen that the accumulation capacitance is equal to the capacitance of the gate dielectric. This means that the k-value for the gate dielectric can be extracted from the accumulation capacitance, C_{acc}, by modelling the capacitor as a parallel plate capacitor as shown below where ε_0 is the permittivity of free space, A_g is the area of the gate and t_{ox} is the thickness of the gate dielectric

$$C_{acc} = \frac{\varepsilon_0 k A_g}{t_{ox}} \tag{3.1}$$

If the dielectric is of different relative permittivity (or k-value) to that of silicon dioxide then it is common practice to calculate the capacitance equivalent thickness (CET) of the dielectric which describes what thickness of silicon dioxide would be required in a MOSCAP device to have the same electrical thickness of the new dielectric. The CET is given by the equation below were ε_{SiO2} is the relative permittivity of silicon dioxide

$$.CET = \frac{\varepsilon_0 \varepsilon_{SiO2} A}{C_{acc}} \tag{3.2}$$

An important parameter for analysing current-voltage plots is the flat-band voltage. To acquire the flat-band voltage, the flat-band capacitance must first be calculated using equation 3.3 below where C_{ox} is the gate oxide capacitance and C_s is the capacitance due to carriers within the silicon substrate.

$$C_{FB} = \frac{C_{ox}C_{s,FB}}{C_{ox} + C_{s,FB}} \tag{3.3}$$

At flat-bands, the value of C_s is given by equation 3.4 [1] below where ε_0 is the permittivity of free space, ε_{si} is the relative permittivity of silicon, A_g is the area of the gate and L_D is the Debye length given by equation 3.5 [1] where k_b is the Boltzmann constant, T is the absolute temperature, q is electron charge and finally N_A and N_D are the carrier doping densities for if the substrate is p-type or n-type respectively.

$$C_{s,FB} = \frac{\varepsilon_0 \varepsilon_{si} A_g}{L_D} \tag{3.4}$$

$$L_D = \sqrt{\frac{k_b T \varepsilon_{si} \varepsilon_0}{q^2 [N_A \text{ or } N_D]}} \tag{3.5}$$

To calculate the Debye length, a figure for the doping density of the silicon substrate must be known. As the tolerances for the doping density of a production wafer are so high then it is better to calculate the value of the doping density from the measured CV characteristics using the "minimum-maximum capacitance method" as shown in equation 3.6 where C_{max} is the measured accumulation capacitance (C_{ox}), C_{min} is the measured inversion capacitance, n_i is the intrinsic doping density and N_D is the doping density of donors for an n-type sample (interchangeable with doping density of acceptors, N_A, for a p-type sample). This equation is transcendental and must be solved numerically.

$$N_D = \frac{4}{A_g^2} \frac{k_b T}{q^2} \frac{1}{\varepsilon_0 \varepsilon_{si}} \left[\frac{1}{C_{min}} - \frac{1}{C_{max}} \right]^{-1} ln\left(\frac{N_D}{n_i}\right) \tag{3.6}$$

Once the value for flat-band capacitance is known then the measured CV plot is consulted to read off the gate bias voltage which causes the measured capacitance to be equal to that of the calculated flat-band capacitance. For an ideal device the flat-band voltage would be zero, however in real devices, the flat-band voltage is modified by parasitic charges as shown in equation 3.7 [1]. Here, the terms are: metal silicon work function difference, W_{ms}, the fixed oxide charge, Q_f, the oxide trapped charge, Q_{ot}, the interface trapped charge, Q_{it}, the mobile oxide charge, Q_m and finally a factor taking account for the charge distribution in the oxide, γ, which is given by equation 3.8 [1] where x is the

distance through the oxide from the gate and insulator interface, t_{ox} is the oxide thickness and $\rho_q(x)$ is the charge density. It can be seen that when the charge is at the gate and insulator interface then it has no effect on the flat-band voltage as it images its charge in the gate.

$$V_{FB} = W_{ms} - \frac{Q_f}{C_{ox}} - \gamma_d \frac{Q_m}{C_{ox}} - \gamma_d \frac{Q_{ot}}{C_{ox}} - \frac{Q_{it}(\phi_s)}{C_{ox}} \tag{3.7}$$

$$\gamma_d = \frac{\int_0^{tox} \left(x/t_{ox}\right) \rho_q(x) dx}{\int_0^{tox} \rho_q(x) dx} \tag{3.8}$$

The fixed oxide charge can be estimated at mid-gap by assuming that the flat-band voltage is caused by only a combination of the metal silicon work function difference and the fixed oxide charge. The metal-silicon work function difference, W_{ms} can be calculated using equation 3.9 [1] below where W_m is the metal work function and W_s is the silicon work function. The metal work function is a "constant" for a material but can be varied by composition, deposition methods and device processing. The silicon work function is given by equation 3.10 [1] where ϕ_f is the Fermi potential and all other variables are as previously defined.

$$W_{ms} = W_m - W_s \tag{3.9}$$

$$W_s = 2\phi_f = 2\frac{k_b T}{q} ln\left(\frac{N_D \text{ or } N_A}{n_i}\right) \tag{3.10}$$

Equation 3.7 can be rearranged to equation 3.11 [1] to calculate the fixed oxide charge assuming negligible trapping effects in the interface and oxide and minimal mobile charge in the oxide.

$$Q_{ox} = (W_{ms} - V_{FB})C_{ox} \tag{3.11}$$

3.2. Current-voltage characterisation (I-V)

The current-voltage plot is a relatively simple measurement technique where a d.c. bias is applied across the MOS capacitor and the leakage current through the device is measured. When current voltage plots are compared between different samples then the flatband voltage calculated above must be taken into account as the charge within the stack affects the effective bias across the stack and so will affect the leakage current magnitude for a given voltage bias on the

gate. Numerous current mechanisms have been observed in high-k dielectrics such as Fowler-Nordheim, Poole-Frenkel, Schottky emission, direct tunnelling and trap assisted tunnelling and these can be identified from the IV characteristics due to their distinctive relationships with the applied field and the temperature. The current mechanisms will not be explored in this thesis.

3.3. Medium energy ion scattering (MEIS)

Medium energy ion scattering (MEIS) is a refined version of the more common technique of Rutherford backscattering spectrometry (RBS) and offers enhanced depth and angle resolution. The experimental setup for the MEIS system can be seen in figure 3.6 below. Light ions (H+ or He+), typically between 100-200 keV, are accelerated and then focused into a precise beam using electrostatic lenses and collimators. The ion beam is made incident on a target specimen along a known direction. The scattered ion beam intensity is analysed for a range of scattering angles and ion energies.

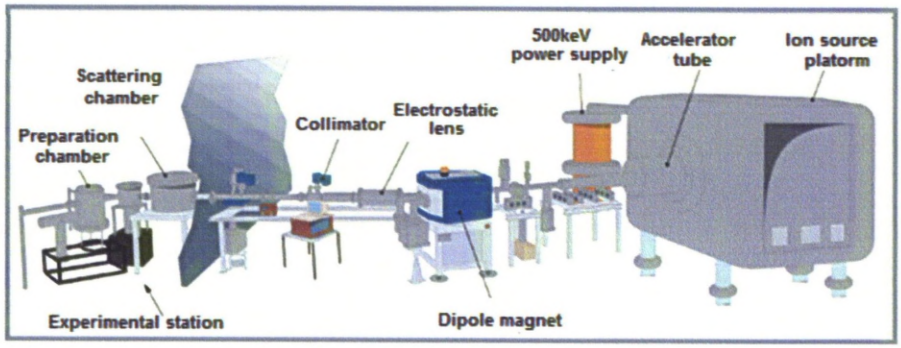

Figure 3.64 - Illustration showing a simplified MEIS setup [2]

The interaction with the sample can produce several effects such as: ion scattering, ion neutralisation and scattering, sputtering, photon and electron emission, adsorption and displacement of specimen atoms/molecules. With MEIS, the main consideration is the scattering of the incident ion beam and more specifically the number of ions scattered by the specimen with specific energies. There are two types of loss which can occur in the scattering event and these are elastic and inelastic as seen in figure 3.7. Elastic collisions cause a large energy loss and the information gained is related to the mass of the target

atom. Inelastic collisions cause a smaller energy loss which is proportional to the length of the path the ion takes through the target material. This allows the extraction of layer thickness with the accuracy of up to one atomic layer. Scattered ions can interact with other atoms within the layer, which can alter the scattering angle as can be seen in figure 3.7. This process is called blocking and can be used to extract information on the material structure. It is common to perform what is called a double alignment before taking MEIS measurements. This term describes a technique were the ion beam is aligned to the sample in a direction so the surface silicon atoms in the substrate of the target shield the atoms in the rest of the layer. This improves the surface sensitivity of the MEIS measurements as the ions only interact with the first few atomic layers of the substrate.

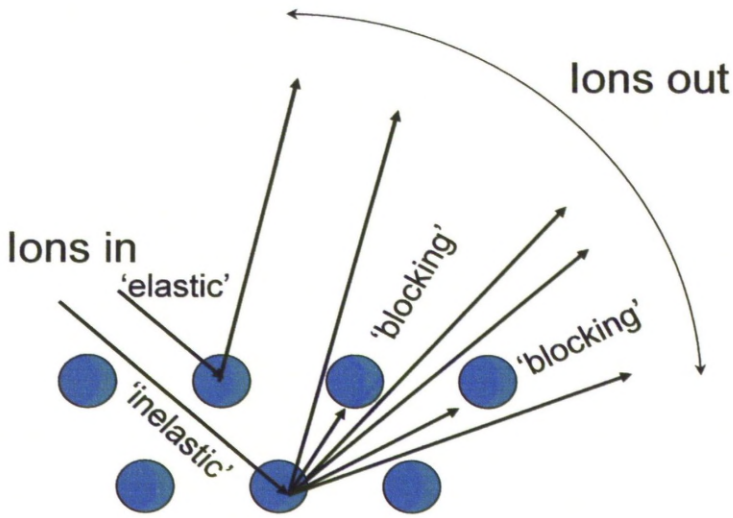

Figure 3.7 - Diagram showing interactions between incident ions and the atoms of the sample material [3]

For the samples for this thesis, the MEIS measurements were performed at the CCLRC Daresbury facility, using a nominally 200 keV He^+ ion beam and an ion dose corresponding to a charge of 10 μC. The samples were aligned to the ion beam along the [$\bar{1}$1$\bar{1}$] channelling direction and the electrostatic energy analyser was positioned to simultaneously record data from along the [111] and [332] blocking directions. These give scattering angles of 70.5° and 60.5° respectively and result in sub-nanometre depth resolution. Analysis was carried out using

the SIMNRA [11] and Igor [12, 13] programs. Using the programs, a model is built up from what is known about the target such as the layer thicknesses, composition and any possible interfaces between the layers. Data is generated for the model accounting for the effects of system resolution and straggling and this is fitted to the experimental MEIS spectra. Quantitative compositional information can be extracted by integrating the area under the peaks on the depth profile and comparing the result to reference data. The scattering cross-section for different elements however, must be taken into account. It is by this technique that the silicon content of the gadolinium silicate layers was extracted in chapter sections 5.3.1 and 5.3.2.

3.4. X-ray diffraction (XRD)

X-ray diffraction (XRD) is a versatile, non-destructive technique that reveals detailed information about the crystallographic structure of a material. It has the advantage that it requires little sample preparation and gives structural information over a whole semiconductor wafer.

There are many types of secondary effects that occur when x-rays interact with matter but fortunately the dominant effect which is measured using a diffractometer, is scattering. In this process, the electron cloud in the path of the X-ray beam vibrates with the frequency of the incoming radiation. The vibration of these charges re-radiates waves of the same photon energy as the incident X-rays by a process called Rayleigh scattering. A crystalline material is a complex, but ordered arrangement of atoms. As an X-ray beam is made incident on this arrangement of atoms, it will be scattered. These scattered X-rays from the different atoms in the structure interfere with one another, cancelling each other out except where the scattered x-rays are "in-phase". In this case the X-rays will scatter constructively to form a new wave. It is this interference which causes the diffraction pattern which can then be studied to discover what crystal phases are present within the specimen.

An ideal crystal arrangement is aligned to diffract monochromatic X-rays of wavelength, λ, from lattice planes spaced at a distance apart of d. The X-rays are incident upon the sample at an angle θ as shown in figure 3.8 below. The

primary beam is partially absorbed and re-radiated from the sample and only the diffracted component is recorded on the detector. The incident angles at which the diffracted beam pattern is at a maximum are given by the equation below, called Bragg's law, where n is an integer describing the order of the diffraction peak and d is the spacing between the atoms within the lattice.

$$\theta_{Bragg} = sin^{-1}\left(n\lambda/_{2d}\right) \tag{3.12}$$

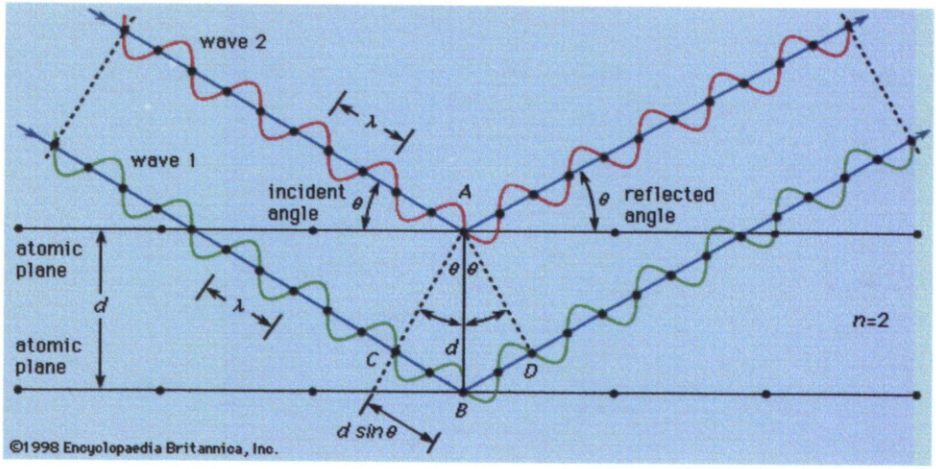

Figure 3.8 - Diagram showing the interference between two x-rays which have interacted with a sample [4]

The angular position of the diffraction peaks produced by an incident x-ray beam of fixed wavelength can be seen to depend on the spacing between the atoms in the lattice. The spectra of a material containing a specific crystalline phase will therefore exhibit peaks at angles which relate to the dimensions of the unit cell associated with this phase. Hence, if the various dimensions of the unit cell or the peak angles caused by a crystalline phase are known then its presence can be identified from the experimental data.

Other than identifying crystal phases which are present within a film, XRD can also provide information on strain in the materials structure. If a crystalline material has a known spacing distance between the atoms in its lattice, d_0, then it will have an easily calculable peak in its XRD spectra as seen in figure 3.9a. If a uniform compressive or tensile strain is applied to the lattice then the distance between the atoms would change resulting in a shift of the characteristic peak in

the XRD spectra as shown in figure 3.9b. Finally, if a non-uniform strain is applied to the crystal lattice as demonstrated in figure 3.9c, then there will be a range of different distances between the atoms in the lattice. This means the diffraction peak will still occur centred at the characteristic angle, but the range of unit cell dimensions will cause the diffraction peak to be much broader.

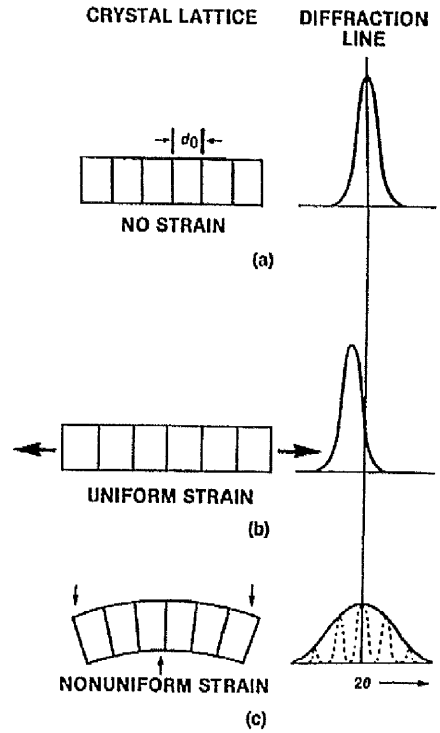

Figure 3.9 - Illustration showing how the XRD diffraction peak of a crystalline lattice can be affected by an applied strain

The XRD measurements displayed in chapter five of this thesis were performed using a Rigaku Miniflex II x-ray diffractometer. This instrument utilises the Bragg-Brentano configuration (also known as θ/2θ) and a diagram illustrating this setup can be seen below in figure 3.10. In this configuration the x-ray source is fixed in position and the angle of incidence of the x-ray beam on the sample, θ_1, is altered by tilting the specimen on a goniometer. For each angle of incidence the angle of the detector is moved to an angle of $2\theta_1$. The detector angle of twice the angle of incidence is no coincidence and is chosen because the x-ray beam formed by constructive interference as discussed earlier is emitted at an angle of twice the incidence angle with respect to the incident

beam. In figure 3.10, r_f is the radius of the focusing circle where the path length from the detector to the sample is the same as the distance from the sample to the source. Also seen in the diagram r_m is the radius of the monochromator focus circle.

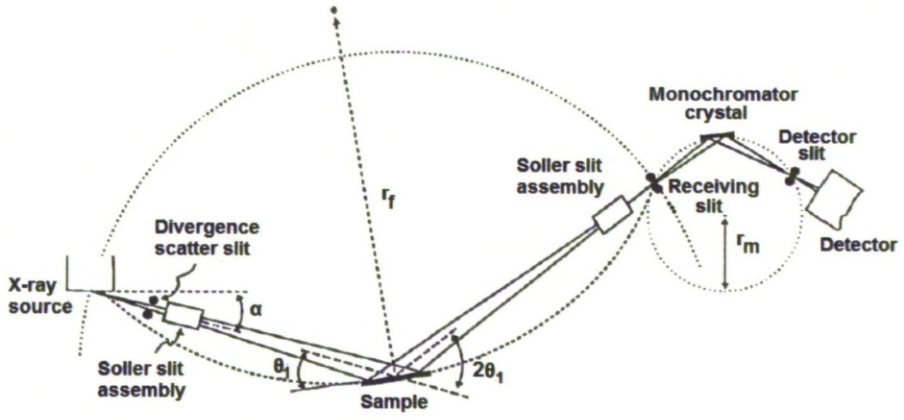

Figure 3.10 - Diagram showing the experimental setup for a Bragg-Brentano XRD [5]

3.5. Transmission electron microscopy (TEM)

Transmission electron microscopy (TEM) is a technique whereby a beam of electrons is transmitted through an ultra thin specimen, interacting with the atoms as it passes through. The interactions of the transmitted form a magnified image onto either photographic film or a sensor such as a CCD. The entire measurement is performed within vacuum.

The resolution of a microscopy system can be evaluated using equation 3.13 [6] and is usually limited by the wavelength of the photons used to probe the sample, λ, and the numerical aperture of the system, NA. The numerical aperture is a dimensionless number that characterises the range of angles over which the system can emit or accept light. It is dependent on the refractive index of the medium, n, and the half angle of the maximum cone of light that can enter or exit the lens, α_{max}, as can be seen in figure 3.11. As the wavelength of light is relatively large (400-700 nm for visible spectrum) then this limited the maximum resolution of optical microscopy

$$.d = \frac{\lambda}{2n\,sin\alpha_{max}} \approx \frac{1.22\lambda}{2(NA)} \tag{3.13}$$

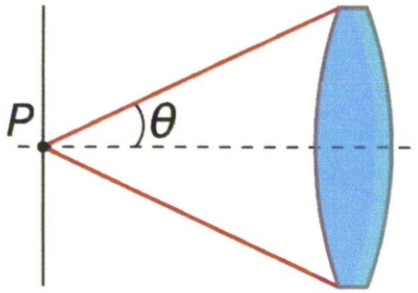

Figure 3.11 - Illustration showing the maximum cone of light which can enter the lens

As electrons exhibit wave-particle duality, a beam of electrons can be considered as electromagnetic radiation. The de Broglie equation gives the relationship between kinetic energy of an electron and its wavelength. As TEM techniques produce electrons close to the speed of light, relativistic effects need to be considered, as shown in equation 3.14 [7] where h is Planck's constant, E is the photon energy of the wave, c is the speed of light in a vacuum and m_0 is electron rest mass.

$$\lambda_e \approx \frac{h}{\sqrt{2m_0E\left(1+\frac{E}{2m_0c^2}\right)}} \tag{3.14}$$

A basic TEM uses multiple components which can be split into four areas; the first is the electron source or electron gun which generates electrons by thermionic emission and accelerates them to high velocities, the second is "lenses" which focus the electron beam before and after transmission through the sample, the third is apertures which restrict the electrons in the incident and resulting (after transmittance through the sample) electron beam to a narrow optic axis and finally a form of imaging system to view the results. An example of a basic TEM system can be seen below in figure 3.12

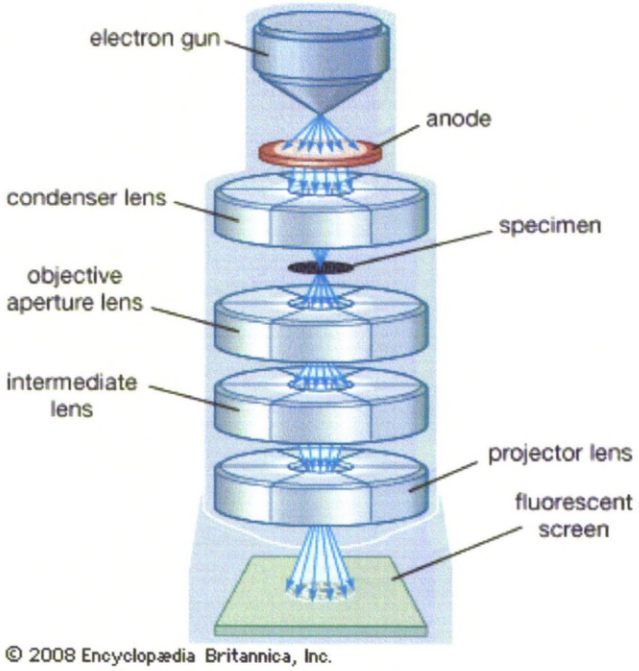

Figure 3.12- Diagram showing a simplified TEM setup [4]

The lenses used are not the same as would be used in an optical system, instead the most common type utilises electromagnetic coils to generate an equivalent to a convex lens. There are typically three stages of lenses which are the condenser lenses, the objective lenses and the projector lenses. The condenser lenses are responsible for forming the initial beam, the objective lenses collate the transmitted beam and the projector lenses focus the electrons that have been transmitted through the specimen onto the imaging device such as a CCD on a modern TEM. The apertures are made up of annular metallic plates which exclude electrons which are travelling too far from the optic axis. The use of apertures removes several problems simultaneously: First they decrease the beam intensity and secondly the filtering of off axial electrons removes electrons that have been scattered by large angles which could be caused by unwanted effects such as spherical or chromatic aberration. The apertures can sometimes be modified to change the aperture size to trade off intensity and the filtering effects.

The TEM instrument produces an image whose intensity can be approximated by the time average amplitude of the electron wave-functions, as given below in equation 3.15 [8] where the wave of the exit beam is denoted by Y, k_b is the Boltzmann constant and t_0 and t_1 signify the start and finish times of the experiment respectively. It can be seen that the observed image depends also on the phase of the electrons.

$$I(x) = \frac{k_b}{t_1 - t_0} \int_{t_0}^{t_1} YY^* \, dt \tag{3.15}$$

A common operating mode, which was used for analysing both the gadolinium oxide samples and the hafnium based samples in this thesis, is the phase contrast mode, which also known as high resolution transmission electron microscopy (HRTEM). The technique relies on analysing the complex modulus of the electron waves exiting the sample which means it is not only the number of electrons which hit the analysing sensor which is considered but also the phase differences of the electron waves. This allows imaging of the crystalline structure on the atomic scale making the technique invaluable for analysing crystallinity, defects, thicknesses and structure. The resolution in a HRTEM is limited by the spherical and chromatic aberration however aberration correction has been implemented in the most cutting edge microscopes called transmission electron aberration-corrected microscopes (TEAM) to give them up to half angstrom resolution [9].

The biggest drawback to TEM is that it is a relatively time consuming process as many materials require extensive sample preparation to reduce the thickness of the sample to dimensions were it is electron transparent. This requires precise use of mechanical milling, chemical etching or ion etching and can potentially inadvertently modify the sample structure.

3.6. X-ray reflection (XRR)

X-ray reflectometry (XRR) is a analytical technique which is used to characterise surfaces, thin films and multilayers. It is related to the complementary techniques of spectroscopic ellipsometry, as discussed in chapter two, and neutron reflectometry. The XRR technique used in chapter

four is called specular scan x-ray reflectometry. In this setup the x-ray source is in a fixed position and the sample is placed on a goniometer which tilts the specimen to alter the angle of incidence. For the scan, the angle of incidence is swept through an angle θ and the detector follows the specular reflected beam sweeping through an angle of 2θ with respect to the transmitted incident beam and a diagram of the experimental setup can be seen in figure 3.13 below.

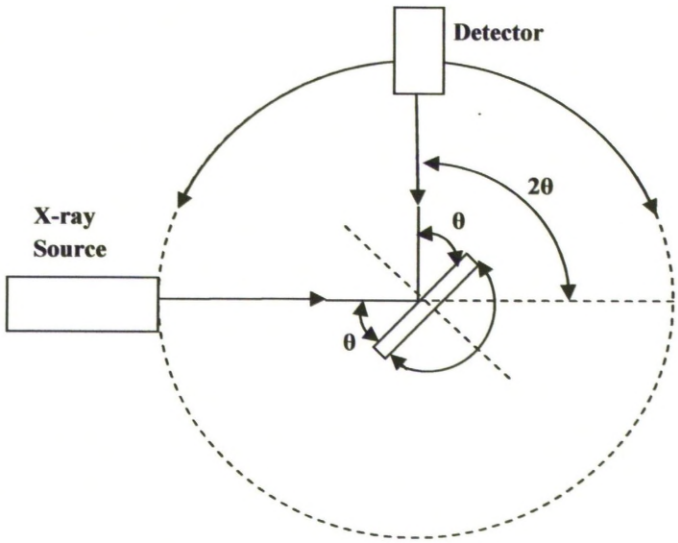

Figure 3.13 - Illustration showing the setup of a spectral x-ray reflectometry measurement

The most important data is measured around the critical angle of external reflection for the sample. Below the critical angle of total external reflection, X-rays penetrate only a few nanometers into the sample. Above this angle the penetration depth increases rapidly. The critical angle for total external reflection is given by equation 3.16 [10] where r_e is the classical radius of an electron, λ is the wavelength for the incident electron wave and N_{av} is the average number of electrons per unit volume. The representation can be changed to photon energy by using equation 3.17 and the value of N_{av} can be replaced by $(\rho/m)n_e$ where ρ is the density of the material, m is the atomic or molecular mass and n_e is the number of electrons in an atom/molecule.

$$\theta_e = \left[\frac{r_e \lambda^2 N_{av}}{\pi}\right]^{1/2} \tag{3.16}$$

$$E = \frac{hc}{\lambda} \tag{3.17}$$

$$\theta_e = \left[\frac{r_e}{\pi}\left(\frac{hc}{E}\right)^2 \frac{\rho}{m} n_e\right]^{1/2} \tag{3.18}$$

Roughness of a sample can be characterised as if the interface is not abrupt, the reflected intensity will deviate from that predicted by the law of Fresnel reflectivity and hence density and roughness can be extracted.

In multilayer stacks, the waves reflected from the different interfaces in the stack interact and hence interference occurs. The interference of the partially reflected X-ray beam causes interference oscillations through the angular sweep which are observed in the reflectivity measurements. An example of an XRR measurement for a three layer stack of a silicon substrate, a ~1nm chemical oxide interfacial layer and a ~4nm hafnium oxide film is shown below.

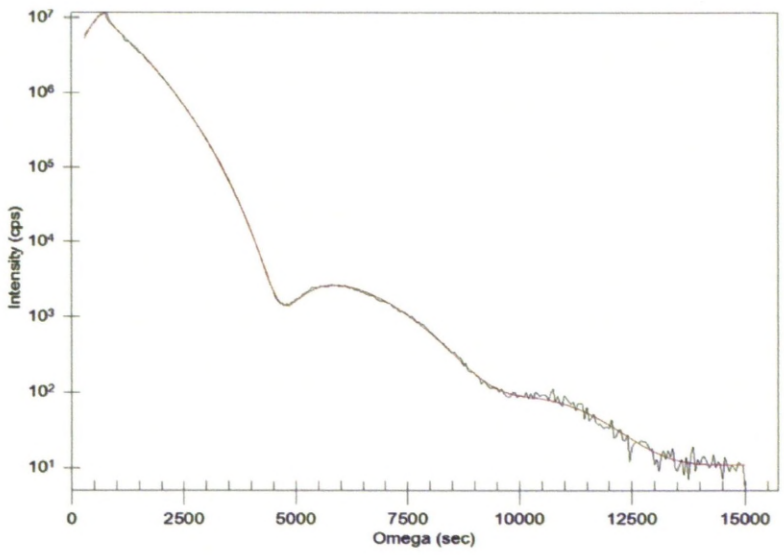

Figure 3.14 - Example XRR measurement for a three layer stack with a silicon substrate, a ~1nm chemical oxide interfacial layer and a ~4nm hafnium oxide

Similar to spectroscopic ellipsometry, the important information about the sample is extracted by creating a model from known information such as the nominal thicknesses of the layers, density and any expected non- idealities such

as roughness and grading. Data is generated for the model and compared to the experimental data. The parameters are then modified by a fitting algorithm until the model accurately represents the measured data. If the fit is insufficiently accurate then the model will be modified by the user including different non-idealities or more layers and the fit process repeated.

3.7. Conclusions

In this chapter, techniques which have been used to analyse high-k gate dielectrics in this thesis have been described and discussed.

3.8. References

[1] D.K. Schroder, "Semiconductor material and device characterization", John Wiley and Sons, (2006), ISBN: 0471739065

[2] CCLRC Daresbury MEIS home page, http://www.dl.ac.uk/MEIS/ (correct address in Sept. 2009)

[3] T. Noakes, Presentation "EPSRC funded 'Mesoscale' facilities at Daresbury", (2008)

[4] Britannica Encyclopaedia online, Www.britannica.com

[5] R. Jenkins and R. Snyder, "XRD - Introduction to X-Ray Powder Diffractometry", Wiley-interscience, (1996), ISBN: 0471513393

[6] Nikon microscopy tutorials, http://www.microscopyu.com (correct address in Sept. 2009)

[7] M.F. L'Annunziata, "Radioactivity: introduction and history", Elsevier, (2007), ISBN: 044452715X

[8] J.M. Cowley, "Diffraction physics", Elsevier Science, (1995), ISBN: 0444822186

[9] Transmission electron aberration-corrected microscopy (TEAM) project home page, http://ncem.lbl.gov/TEAM-project/index.html, (Correct address in Sept. 2009)

[10] Y. Chung, "Practical guide to surface science and spectroscopy", Academic Press, 2001, ISBN: 0121746100

[11] SIMNRA Program, M. Mayer, Max Planck Institute of Physics, http://www.rzg.mpg.de/~mam/ (correct address in Sept 2009)

[12] Igor pro version 6.1, Wavemetrics, http://www.wavemetrics.com/ (Correct address in Sept. 2009)

[13] Custom macros for MEIS data analysis (designed for Igor Pro), Daresbury Lab, http://www.dl.ac.uk/MEIS/software/index.htm (Sept. 2009)

Chapter 4

Hafnium oxide and hafnium silicate films

4. Hafnium oxide and hafnium silicate films

4.1. Introduction

The chapter commences with a literature review of hafnium based dielectrics. This is followed by experimental results on hafnium oxide and hafnium silicate films. The physical and electrical characteristics of the layers are measured to allow a study into the effect of altering the silicon content of a hafnium silicate film. New approaches are used for characterization of metal oxides (high-k) thin films using SE, including density estimation for hafnium oxide and hafnium silicate films; the density is an important physical characteristic of a film and is required for the interpretation of assessment of medium energy ion scattering results. Finally, a technique is described, using SE and Capacitance Voltage (C-V) measurements, for de-convoluting the lattice contribution to the relative permittivity and its dependence on the film composition. An anomalous shape in the graph of hafnium content versus lattice contribution is explained by further analysis using MEIS and TEM and the three techniques are compared as a thickness measurement.

4.2. Literature review for hafnium oxides and silicates

Hafnium oxide was one of the first dielectrics to show promise as a possible high-k replacement material for silicon dioxide. However, in practice, deposited hafnium oxide gate stacks suffer many problems such as degraded field effect carrier mobility [1], Fermi-level pinning at the poly-silicon/HfO_2 interface [2], threshold voltage instability during processing [3] and also during device operation [4]. The Fermi-level pinning problem at the poly-silicon/HfO_2 interface is one of the main reasons why metal gates are preferred for HfO_2 devices. Instability during high temperature processing [5] is also a major problem as the film becomes polycrystalline with grain boundaries which are known to enhance the diffusion of dopants through the film [6, 7] and also increase film leakage. Another major problem is the formation of a low-k layer at the film/silicon interface which lowers EOT of the dielectric stack. There are three proposed solutions to these problems which are: a) the incorporation of nitrogen into the film [8], b) the use of hafnium silicates [9] or c) the use of hafnium aluminates [10].

Hafnium silicates are grown by incorporating silicon into a hafnium oxide layer during deposition. The expected advantage to using hafnium silicates compared to the pure oxide is that this should produce a film with increased crystallisation temperature [11], which has a stable amorphous structure [6, 12, 13] and provides resistance to oxygen diffusion [6,11] , thus avoiding the formation of any interfacial layers. They however suffer from a lower k value than the pure oxide having a k value measured to be k=11-12.4 [14] (compared to 21-25 for HfO_2 [15, 16]) from Capacitance Voltage (C-V) analysis.

Fowler-Nordheim, Poole-Frenkel and trap-assisted tunnelling leakage mechanisms have been observed in hafnium oxide and hafnium silicate films and from associated analysis the conduction band offset for hafnium oxide has been reported to be 1.4eV [14] which is very close to the theoretically calculated value for conduction and valence band offsets which are 1.5eV and 3.4eV respectively [17]. Xu et al [18] examined the leakage currents through a hafnium oxide film under different bias regimes. It was discovered that at low oxide fields under both gate and substrate injection regimes the shallow traps in hafnium oxide could not contribute to leakage and so only direct tunnelling and trap assisted tunnelling were observed. At higher oxide fields there was seen to be a transition for both gate and substrate injection to Fowler-Nordheim and Poole-Frenkel mechanisms with Poole-Frenkel hopping dominating leakage at the highest bias voltages. The shallow trap depth was extracted from the Poole-Frenkel plot and values of trap depth were found to be between 0.5-0.8eV below the hafnium oxide conduction band. From Poole-Frenkel analysis trapping levels in hafnium silicate have been observed to occur deeper than hafnium oxides at 1-1.1eV under the conduction band edge [19]. During the P-F analysis it was possible to extract the dynamic permittivity from the I-V characteristics which is reported to be 3.2-3.5 [20].

Takeuchi [20] combined results from many published studies on hafnium silicates to establish the relationship between the relative permittivity of a hafnium silicate film and its hafnium oxide content. Takeuchi commented that due to the lack of data for similar film preparation methods, the variation between the various experimental data was too large to formulate a consistent

trend. To solve this problem, Takeuchi proposed the use of three separate relationships shown in figure 4.1 below as the three possible simple trends.

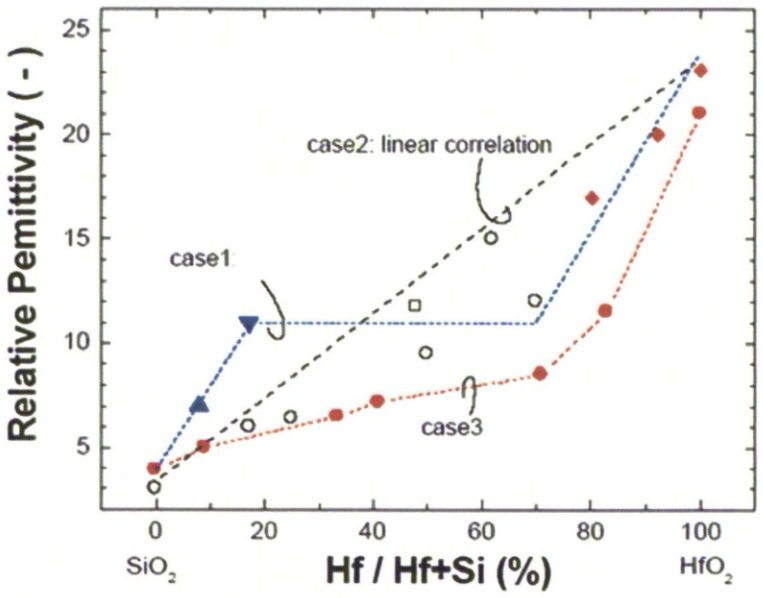

Figure 4.1 - Plot showing proposed trends for current experimental relative permittivity versus film hafnium content for hafnium silicate films [20]

It can be seen from figure 4.1 that, as expected, for increasing hafnium content of the film, the relative permittivity of the film increases. However from the plot it is noticeable that there is unlikely to be a linear correlation found to explain the relationship and hence the case 3 trend was included to provide a reference by which to observe the other trends.

A relationship between the optical band-gap of a hafnium silicate film and its hafnium oxide content has also recently been reported which shows the band-gap decreases by roughly 50meV for hafnium oxide up to 64% at which the band-gap becomes constant at 5.7eV as shown in figure 4.2 below. It is expected that the reason for lower leakage of hafnium silicate compared to hafnium oxide is this greater band-gap and also greater effective electron rest mass of hafnium silicate [21].

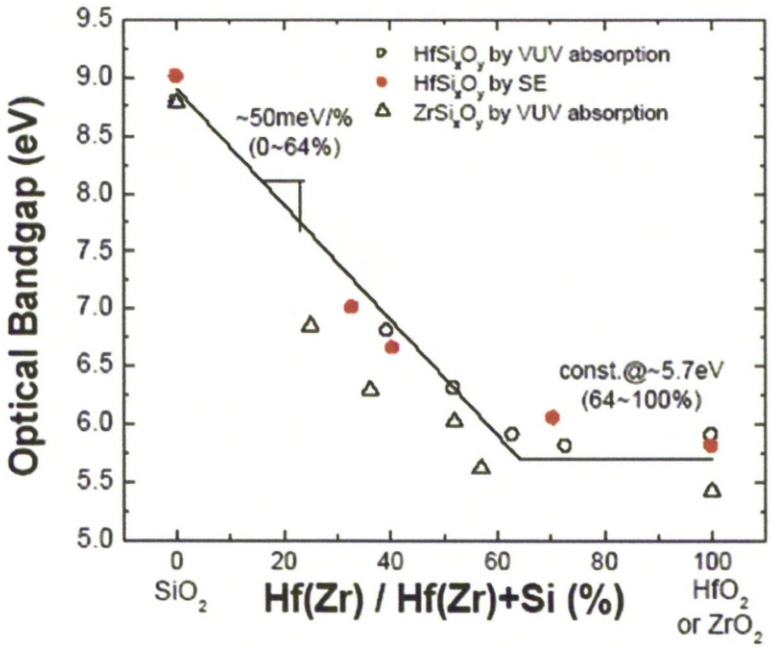

Figure 4.2 - Graph showing decrease of optical band-gap of a hafnium silicate film with increasing hafnium content (Please see Takeauchi et al for references) [20]

The structure of hafnium silicates is expected to be similar to that of zirconia because zirconium and hafnium share the same group in the periodic table, and hence the same outer electron configuration. Wilk et al [22] published XPS results for sputtered hafnium silicate layers which showed Hf-O and Si-O bonding within the film, but no Hf-Si features were present. Lucovsky et al [23] have proposed that bonding constraints at the silicon/insulator interface had to be considered because systems that are over- or under-constrained with respect to SiO_2, possess a high density of electrical defects near the silicon/dielectric interface and this is known to result in poor electrical properties. Wilk et al [22] explained in their review of high-k dielectrics that this meant that any silicide bonding which forms near the channel interface will tend to form unfavourable bonding conditions leading to increased leakage current and reduced electron channel mobility [24]. High leakage currents associated with hafnium silicide have been observed experimentally by oxidising pure hafnium silicide layers to produce hafnium silicate films [25]. The effect of sputtering temperature on the characteristics of hafnium silicate layers has been investigated by XPS [22] and it was observed that higher sputtering temperatures produced more Hf-O and Si-

O bonds while avoiding silicide formation. Hafnium silicate films of similar composition produced by e-beam evaporation at a temperature of 600°C exhibited higher levels of Hf-O and Si-O bonding within the film, however, there were also peaks present corresponding to silicide bonding [22]. Due to the silicide formation it would be expected that transistors incorporating e-beam evaporated material would have inferior electrical characteristics to those with sputtered gates. Sputtered hafnium silicate films containing low hafnium content and capped with amorphous silicon have been observed to remain amorphous and stable for a poly-silicon gate activation anneal of 1050°C for 20s in pure nitrogen. It was proposed that hafnium silicates with a hafnium content of less than 30% would remain amorphous at process annealing temperatures [22].

Cho et al [26] studied phase separation of hafnium silicate as a function of composition and temperature by thermally annealing ALD deposited hafnium silicate, $(HfO_2)_x(SiO_2)_{1-x}$, samples of compositions x =0.25, 0.50 and 0.75 in N_2 at temperatures above 800°C. Cho reported that the degree of the sharpening of the HfO 4f peak in the XPS peak spectra was related to the degree of phase separation within a hafnium silicate film and from studying the spectra he was able to ascertain that phase separation does not occur even at temperatures of 900°C for a x=0.25 sample, however the x=0.75 sample cannot resist phase separation at only 800°C [26]. Cho proposed from MEIS measurements taken of annealed samples that during annealing, small SiO_2 grains intrude into the $HfSiO_2$ layer at the high-k film surface but as the temperature is increased up to 900°C the quantity of extracted SiO_2 at the film surface exceeded that of the incorporated mole fraction of SiO_2 in the as-grown film [26]. This meant there was another source causing the remaining SiO_2 at the high-k layer surface and this was reported be the result of diffusion from the Si substrate through the HfO_2 film.

The inclusion of nitrogen into the film refers to the deposition of a hafnium silicate film in an environment containing nitrogen gas or post deposition annealing in an environment which would diffuse nitrogen into the film. Both these methods lead to the formation of a hafnium silicon oxynitride film (HfSiON) which was studied by M. Koyama et al [8]. This study showed that

the inclusion of nitrogen reduced the formation of crystalline grains of HfO_2 within the film during poly-silicon activation annealing (N_2 at 1100C for 15s) and the film was shown to be completely homogeneous when the nitrogen concentration was above a concentration of 30 at %. As well as resisting phase separation, the films were seen to show extremely stable k-values even after annealing at 1000°C in pure nitrogen. Analysis of the boron concentration as a function of depth in the samples showed a considerable reduction of the boron penetration depth into the silicon substrate as the nitrogen concentration within the film was increased. Inclusion of nitrogen into a hafnium silicate film however has been reported to reduce the band-gap and band offsets of the silicon/HfSiON interface [27]. By comparing a hafnium silicate layer containing 40% HfO_2 which had been nitrided, to an as-deposited sample it was observed that the nitridation of the HfSiO film reduces the band-gap of the film by ~1.50eV compared to a pure hafnium silicate film. From further analysis it was discovered that the valence band and conduction band offsets with the silicon substrate had been reduced by ~1.2eV and 0.33 eV respectively when compared to pure hafnium silicate. This reduction in band offsets are reported to not be serious enough to affect the viability of HfSiON for use in future CMOS applications, however they will cause an increase in the leakage currents measured through such a film.

Cho et al [28] compared the effects of pure nitrogen and pure ammonia annealing on hafnium silicate films containing a high hafnium mole fraction of 0.75. Hafnium silicate films of thickness ~3.5nm were deposited by ALD on top of silicon substrates which were covered by a ~1 nm thermal SiO_2 layer [28]. The samples were split into two groups which were exposed to RTA annealing in either N_2 at a temperature of 900°C for 1 min or NH_3 at a temperature of 900°C for 1 min. It was reported that monoclinic crystal grains of hafnium oxide were formed within the hafnium silicate film which was annealed in pure N_2, however no such crystal regions were reported for the film annealed in NH_3 and this suggests that NH_3 might suppress the phase separation of hafnium silicate films. The annealing process however also increased interface layer of both samples by approximately 2nm and reduced the thickness of the layers with the thickness reduction of the high-k layer being almost four

times bigger in the sample annealed in N_2. Near edge x-ray absorption fine structure (NEXAFS) spectra measured on the sample annealed in pure nitrogen exhibited peaks associated with phase separation; however, the sample annealed in NH_3 was observed to be amorphous and homogenous indicating that the NH_3 ambient aids the resistance of phase separation. The suppression of phase separation was expected to be caused by nitrogen incorporation as annealing in NH_3 incorporated a nitrogen concentration of ~20%. The nitrogen was seen to be uniform with depth and the oxygen concentration of the film was seen to fall in parallel with the increase in nitrogen concentration. This suggests that the nitrogen substitutes for oxygen within the material and this was confirmed by XPS analysis by H-J Cho et al [29] which reported a decrease in Si-O bonds as the concentration of Si-N bonds increased during NH_3 annealing. Electrical characterisation of MOS capacitors formed with platinum electrodes was carried out and a k-value of ~15 was extracted from HFCV measurements [28]. It was observed that the flat-band voltage shift for samples annealed in NH_3 was negative suggesting that there was an increase in positive oxide charge, presumably due to the nitrogen incorporation. The flat-band voltage shift for the samples annealed in N_2 was seen to be positive which suggested either a decrease in positive oxide charge due to an improved interface by out-diffusion of impurities, an increased interface layer thickness or increase in negative trap charge which could be caused by an increase in grain boundaries due to phase separation.

H-J Cho et al [29] also compared the effects of annealing of ALD prepared hafnium silicate films in NH_3 or N_2. MOS transistors were formed using poly-silicon gates from films either annealed in NH_3 at a temperature of 750°C for 60s or annealed in N_2 at a temperature of 950°C for 60s. The EOT of the samples annealed in pure nitrogen and NH_3 was only 0.1nm higher and 0.25nm smaller respectively than that of the as-deposited film. The normalised saturated drain current of the MOSFETs were measured and it was seen that annealing in a pure nitrogen environment causes an increase of 24% compared to the as-deposited for a corrected bias voltage of 1.2V (V_g-V_{th}). The NH_3 anneal however degrades the normalised drain current by 4% compared to the as-deposited gates at a corrected bias voltage of 1.2V. This is at least partially

due to the degradation in the effective mobility of the carriers possibly resulting from a large nitrogen concentration at the gate dielectric/channel interface, as discussed earlier. As expected the effective carrier mobility in the associated MOS transistors annealed in pure nitrogen was over 12% larger than that of both the as-deposited and NH_3 annealed MOS transistors. The best compromise between the film stability and effective mobility was to first anneal in NH_3 at 750°C for 60s followed by the pure nitrogen anneal at 950°C for 60s. This sequence incorporated nitrogen into the film to prevent phase separation whilst the nitrogen anneal improved the effective mobility of the carriers to that of the MOSFETs which had been annealed only in the pure nitrogen. The pure nitrogen anneal was seen to only remove ~1.5% of the nitrogen incorporated by the NH_3 anneal (13%).

Finally, it is worth noting that hafnia based dielectrics have been successfully incorporate into production by Intel [30].

4.3 Experimental results

4.3.1. Sample preparation

Hafnium oxide films with nominal thicknesses of 4nm, 6nm and 8nm were deposited by atomic layer deposition (ALD) on p-type Si wafers covered with a nominally 1nm thick rapid thermal oxide (RTO) SiO_2. Also hafnium silicate, $(HfO_2)_x(SiO_2)_{1-x}$, thin films (x=0.3, 0.5, 0.7 and 1.0) of 4nm nominal thickness were deposited using the metal organic chemical vapour deposition (MOCVD) process, upon p-type silicon wafers coated with, a chemical oxide of ~ 1 nm nominal thickness.

4.3.2. Spectroellipsometric measurements and data analysis methodology

Two different J. A. Woollam variable angle spectroscopic ellipsometers were used for analysing the samples. The thickness of the samples was estimated using a J.A. Woollam M2000UI spectroscopic ellipsometer which can record data for a series angles over a spectral range of 240-1700nm (5.2-0.7eV). Measurements were performed for three angles of incidence (65, 70 and 75) across the measurable spectral range. Further measurements on the band gaps and band gap variations were performed by collaborators at J.A. Woollam inc. using a J.A. Woollam VUV-VASE spectroscopic ellipsometer, which can also

measure data for a number of angles but covers a spectral range of 140 nm to 1700 nm (8.87 – 0.73eV respectively) using a nitrogen purge to avoid the atmospheric absorptions of UV radiation. The VUV-VASE measurements were performed at multiple angles of incidence (70° - 75°, 1° steps) over the measurable spectral range and data acquisition and analysis was performed with WVASE version 3.517.

4.4. Analysis of spectroscopic data

For each of the different interfacial films, there was a bare control sample grown under the same conditions as the interfacial film used in the samples deposited with high-k layers. This allowed analysis of the interfacial layer alone, this reducing the complexity of analysing a full high-k stack into two much simpler stages. It was assumed that the optical properties of the interfacial layer were not significantly altered by the deposition of the hi-k material. The two models used in the fitting can be seen below in figure 4.3.

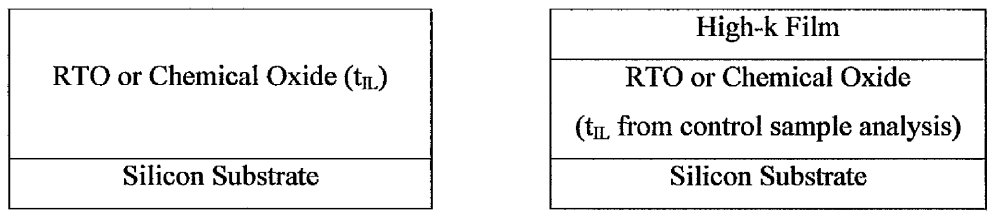

Figure 4.3 - Two models showing the procedure for fitting the hafnium oxide/silicate samples. Model a) was used to analyse the interfacial layers and then the properties of the interfacial layer was fixed in model b) to extract the properties of the high-k layer

The rapid thermal oxide (RTO) interfacial layer was measured using spectroscopic ellipsometer and fitting was performed using the database refractive index values for a thermal oxide layer. To improve the accuracy of the fit, a Cauchy layer fit procedure was carried out as explained in Chapter 2.10 fitting the film thickness and first Cauchy parameter in the wavelength region 600-1700nm and then fitting the three Cauchy parameters over the entire measured spectral range.

The chemical oxide interfacial layer was measured using spectroscopic ellipsometer and fitting was attempted with both a database silicon dioxide

layer, fitting for thickness alone, and a Cauchy layer where the fitting was performed for thickness and either 2 or 3 Cauchy parameters, in the transparent region. These layers gave the expected values for thickness. However, when expanding the fitted spectrum over the entire measured range, it proved impossible to fit to the experimental data to produce acceptable physical results for the optical constants. The optical constants were therefore fitted using a Lorentz layer with two oscillators. This Lorentz interfacial layer was used when analysing the high-k films. The chemical oxide thickness extracted from this control sample t_{IL}=1.6nm, was kept constant in all further models.

The thickness for the HfO_2 and $(HfO_2)_x(SiO_2)_{1-x}$ layers was then extracted using a Cauchy layer in the transparent region. The results obtained for the thicknesses of the high-k layers, are summarised in the table below:

Sample molecular HfO_2 content, x	**High-k film thickness (nm)**
0.3	2.41
0.5	3.19
0.7	3.31
1.0	4.29

Table 4.1 - Table summarising the film thicknesses extracted for the hafnium oxide and hafnium silicate thin films using SE

The thicknesses of the layers were now fixed and a point-by-point fit was performed over the entire measured range, to allow extraction of the optical constants. The imaginary component was first fitted alone to the prototype point-by-point extracted optical constants using two Tauc-Lorentz oscillators. The Sellmeier poles were then fitted in standard order with the amplitude of the highest energy pole fitted first, followed by then the energy of the high energy pole and then the amplitude of the long wavelength pole. The real part of the permittivity, ε_1 was then fitted to improve the fit to the point-by-point prototype layers. A fit was then performed to the experimental data to check the accuracy of the oscillator model. The optical constants for the various hafnium oxide and silicate layers which were extracted using this procedure are shown in figures 4.4 and 4.5. It can be seen in figure 4.4 that the higher the molecular content of

silicon, then the lower is the real value of the relative permittivity at the lowest energy. It can be seen also that as the silicon dioxide content increases, then the peak seen in the real part of the relative permittivity decreases significantly.

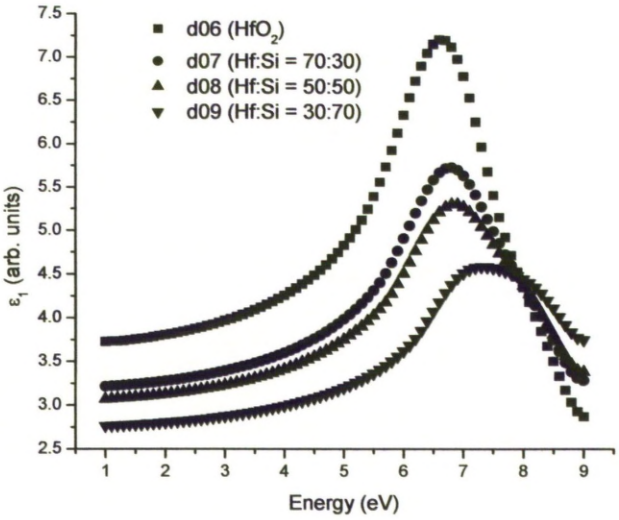

Figure 4.4 - Real dielectric constant extracted using general oscillator model for MOCVD hafnium silicate thin films

Turning to figure 4.5, it can be seen that the imaginary part of the relative permittivity, ε_2 at the band edge decreases in magnitude as the silicon dioxide content increases. The maximum of the absorption on the plot also looks to shift to higher energies with increasing silicon dioxide content. From the imaginary part of the optical constants in figure 4.5, the absorption coefficients, α, were calculated using the technique in chapter two and by using a Tauc plot of $(\alpha nE)^{1/2}$ versus E, both the direct and defect related band gaps were extracted. The plots are shown below in figures 4.6-4.9.

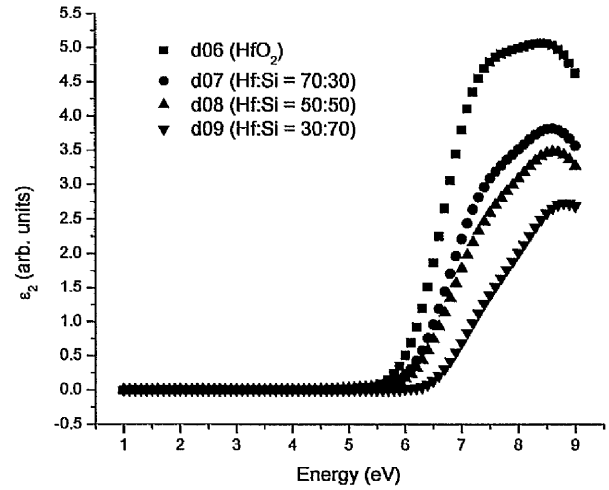

Figure 4.5 - Imaginary dielectric constant extracted using general oscillator model for MOCVD hafnium silicate thin films

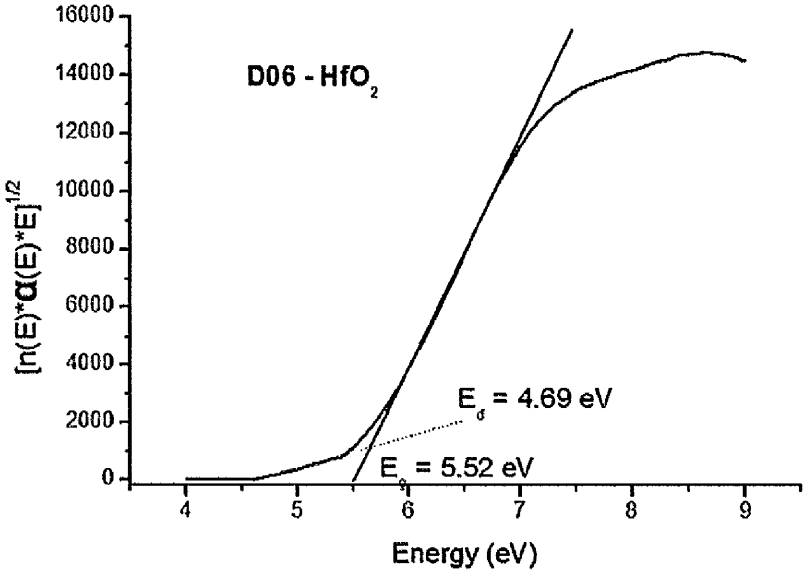

Figure 4.6 - Tauc plot for Sample D06 (100% HfO_2) showing band gap and near band edge state

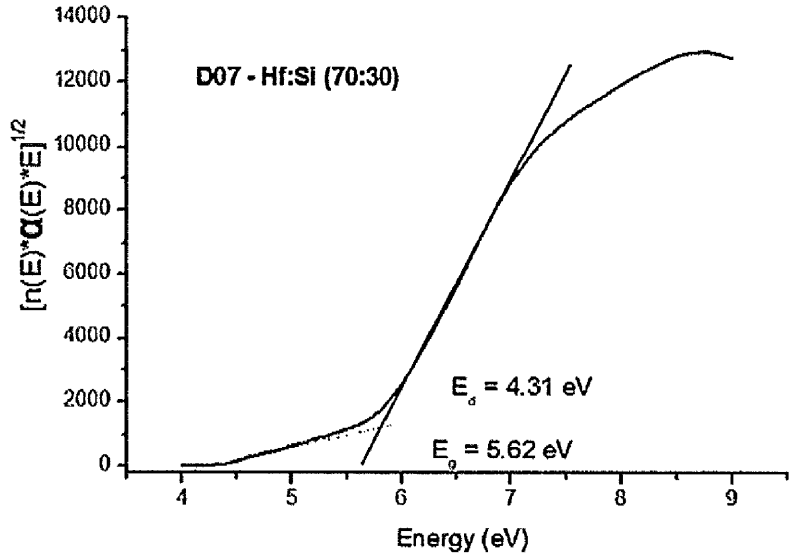

Figure 4.7 - Tauc plot for Sample D07 (70% HfO_2) showing band gap and near band edge state

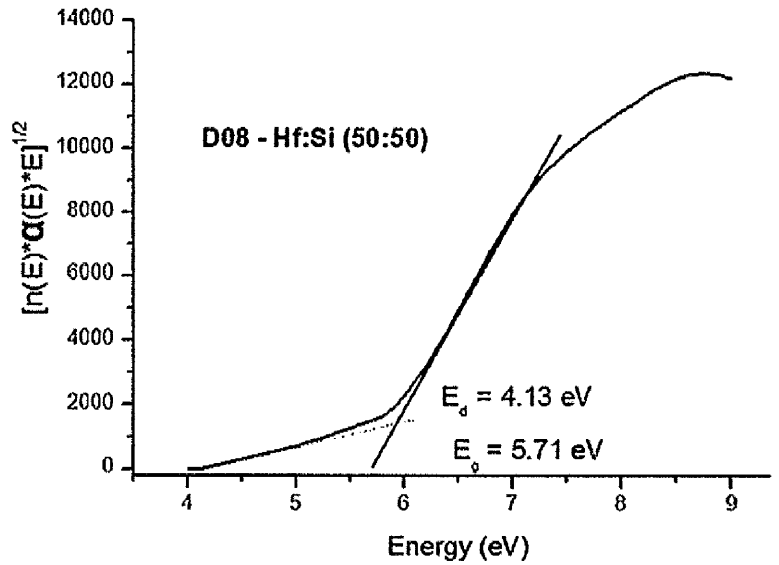

Figure 4.8 - Tauc plot for Sample D08 (50% HfO_2) showing band gap and near band edge state

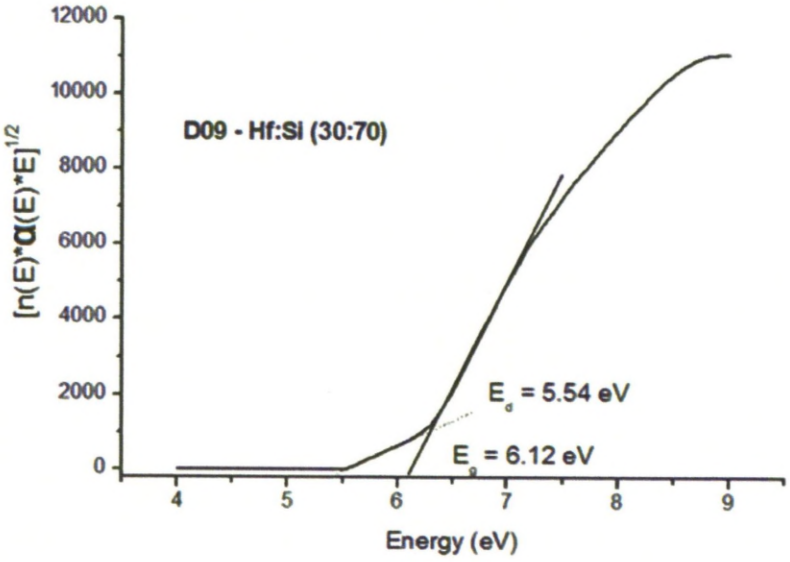

Figure 4.9 - Tauc plot for Sample D09 (30%HfO_2) showing band gap and near band edge state

The band gap results are summarised in table 4.2 below

HfO_2 molecular content, x $[(HfO_2)_x(SiO_2)_{1-x}]$	Band gap (eV)	"Near" band gap defect level (eV)	$\Delta E_g = E_g - E_{defect}$ (eV)
1.0	5.52	4.69	0.83
0.70	5.62	4.31	1.31
0.50	5.71	4.13	1.58
0.30	6.12	5.54	0.58

Table 4.2 - Summary of the band gap values and near band edge defect levels for the hafnium oxide and hafnium silicate films

The energy position of the band gap and the 'near' band gap defect or energy level, were plotted against the molecular content of hafnium oxide as shown in figure 4.10. The band gap is seen to decrease as expected, with increasing hafnium oxide molecular content. The largest increase is between a hafnium oxide molecular content of 0.3 and 0.5. The near band gap defect level is relatively shallow in pure hafnium oxide but almost doubles in depth as the

silicon dioxide content of the high-k layer increases. This effect has been reported elsewhere [19]. The depth of this energy level for the hafnium silicate layer with the lowest silicon dioxide content shows an interesting result because it is lower than that for pure hafnium oxide. The reason for this is unknown but it may be an erroneous result for the near band gap energy level as the trend for this sample changes significantly compared to the other higher hafnium oxide content samples. However it could be related to the layer now behaving more like a silicon dioxide with hafnium oxide impurities as opposed to a hafnium oxide with silicon dioxide impurities. It can be seen that the energy level of the defect in the lowest hafnium content layer is the same as the band-gap of pure hafnium oxide. Therefore it is possible that the hafnium oxide impurities in the mainly silicon dioxide layer is now providing an energy level in the band-gap of the film.

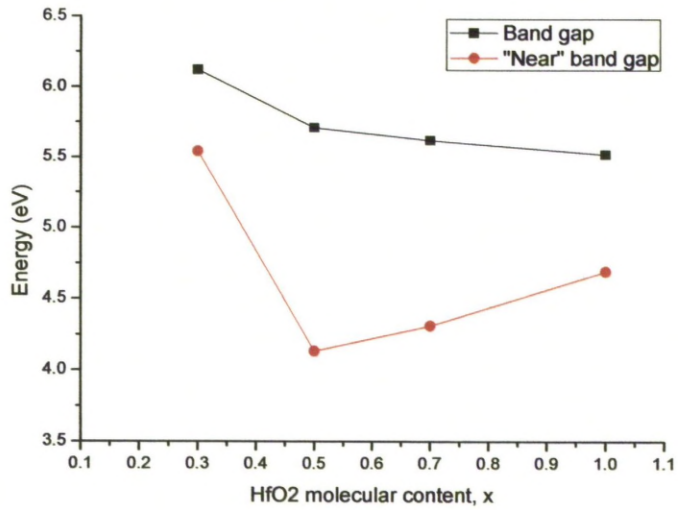

Figure 4.10 - Graph showing the band gap and the near band gap energy level or defect plotted versus the hafnium oxide molecular content

4.5. Electrical analysis of the hafnium oxide and hafnium silicate samples

After ellipsometry measurements had been carried out, a sample of each wafer was taken and MOS capacitors were formed by depositing gold as the gate electrode on using a shadow mask arrangement to define large, circular capacitors with area 7.1×10^{-8} m^2 and 4.9×10^{-8} m^2.

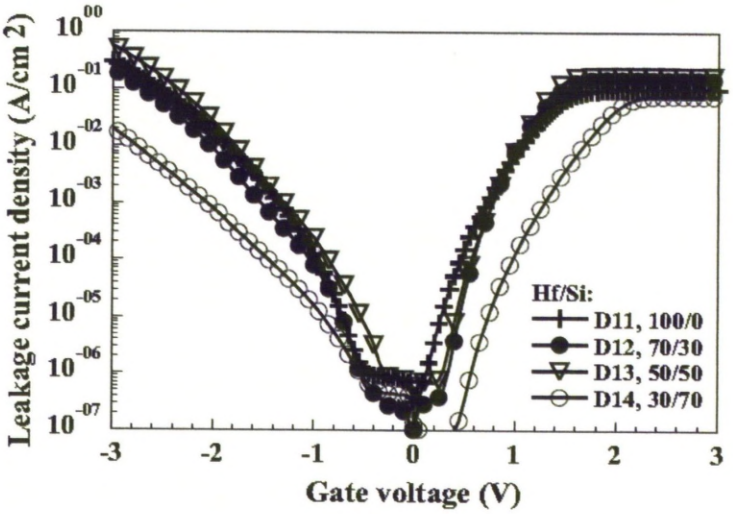

Figure 4.11- Graph showing the leakage current densities measured for the hafnium silicate thin films of different hafnium molecular content for both the positive (substrate injection) and the negative gate bias (gate injection) regimes

Current-voltage measurements were taken for all of the samples in both the negative and the positive regimes and the resulting current density versus gate voltage plots can be seen in figure 4.11. It can be seen that the lowest leakage currents (by over an order of magnitude) were measured for the layer which had the lowest hafnium molecular content. If the current is barrier limited, this would be expected as these layers had the highest band-gap and band offsets, based on the ellipsometry results. The highest leakage was measured for the layers with a hafnium oxide content of ~50%. A more detailed study of leakage current measurements is required to resolve the latter anomaly but this is beyond the scope of the study here. The leakage current in the substrate injection regime is seen to saturate. This saturation is caused by a lack of supply of electron minority carriers associated with the p-type substrate, in the inversion operating regime.

C-V measurements were taken across a range of frequencies between 1 kHz and 1 MHz to reveal large frequency dispersion. This was likely to be related to a 'leaky interface' and the correction technique proposed by Yang et al [31] was applied to extract the true accumulation capacitance of the stack. The capacitance plots for all of the samples are shown in figure 4.12. The extracted parameters for the devices are shown in table 4.3 and the hafnium silicate

(except x=0.7) films can be seen to display slightly lower flat-band voltages of ~0.1V compared to the hafnium oxide sample. This was reported in [32], where it was suggested that it could be due to the role of hydrogen in the films. Interstitial hydrogen has been reported to act as a shallow donor in hafnium oxide films, but it is suggested that this defect lies deeper in the band gap of hafnium silicates. This could also explain the near band edge defect seen in the ellipsometry results.

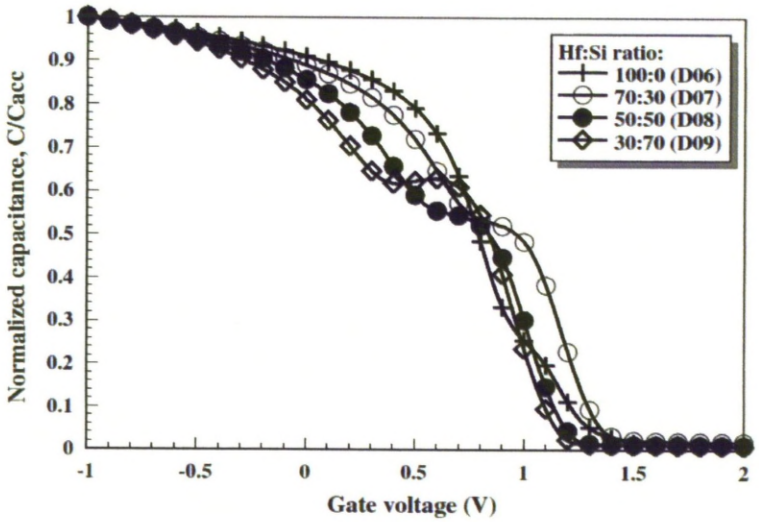

Figure 4.12 - Frequency corrected CV plots for $(HfO_2)_x(SiO_2)_{1-x}$ samples containing various levels of hafnium oxide molecular content (Cacc: D06=711.7pF, D07=925.6pF, D08=627pF and D09=600.7pF) (Area: D06, D08 and D09 = $4.9x10^{-8}m^2$ and D07 = $7.1x10^{-8}m^2$)

Sample	HfO_2:SiO_2	SE high-k (nm)	SE interfacial layer (nm)	Effective oxide thickness (nm)	Flat band voltage (V)	Fixed charge density (10^{12})(cm^{-2})
D06	100:0	4.3	1.6	2.38	1.27	-10.4
D07	70:30	3.3	1.6	2.64	1.28	-8.1
D08	50:50	3.2	1.6	2.70	1.17	-6.8
D09	30:70	2.4	1.6	2.82	1.14	-6.5

Table 4.3 - Structural and electrical data extracted for the hafnium oxide and hafnium silicate layers with various hafnium oxide content

The oxide fixed charge was calculated using the work function of 5.1eV for the gold gate material. The charge was found to be negative in polarity and of the order $>5x10^{12}cm^{-2}$. This value is large; however the fixed oxide charge is seen to decrease with increasing molecular content of silicon dioxide. The dielectric constant was extracted from the C-V data using the SE measured thicknesses and the pure hafnium oxide was found to have a dielectric constant of 21 which reduces to a k-value of ~8 for the hafnium silicate with a molecular content of 0.3. The dielectric constant was plotted versus molecular hafnium oxide content for these samples along with other published results to allow comparison and this can be seen in figure 4.13. The dielectric constant results for these samples are seen to be similar to previously published work, but appear to have slightly higher values than other MOCVD deposited hafnium silicate samples. As reported previously by Takeuchi [19] for hafnium silicate, the k-value seen in figure 4.13 is seen to vary non-linearly with hafnium oxide content and so fitting with a single curve is difficult. To aid viewing of the trends a line was drawn between the results.

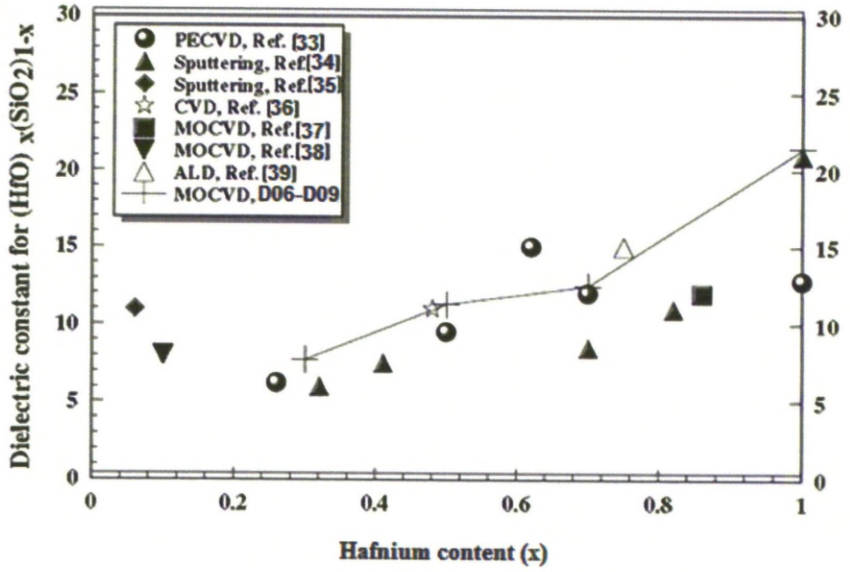

Figure 4.13 - Graph showing the relative dielectric constant for these hafnium oxide and hafnium silicate samples along with results reported by other research groups

Much work has been published on the prediction of the optimum composition of hafnium silicate film. Callegari [34] proposed that the optimum composition

would have to be less than ~30% hafnium oxide content to avoid the crystallisation during high temperature processing. Takeuchi proposed a figure of merit $k(m^*\Phi_B)^{0.5}$ which included the effective tunnelling mass, m*, the relative permittivity , k, and the barrier height, Φ_B. Using this model, Takeuchi predicted that the optimum hafnium content was ~20%. However it can be seen from figure 4.13 that, such low hafnium content will provide k-values of less than 8 which is unacceptable for a gate dielectric replacement material. Krug et al [40] also acknowledged this problem and proposed an optimum hafnium oxide content of 60% which would give a k-value of ~12 from figure 4.13.

4.6. Density extraction using optical techniques

4.6.1. New technique for extraction of density of hafnium oxides

The relation between the density of inorganic solids and their optical properties is well established [41] and given by the Clausius-Mossotti shown in equation 4.1. A simple relation between the refractive index (n) and density (ρ) is apparent and so the density of a hafnium oxide film can be extracted, using the refractive index obtained from SE:

$$\frac{n^2-1}{n^2+2}=\left(\frac{4\pi}{3V_m}\right)\alpha_e=4.19\left(\frac{\alpha_e}{V_m}\right)=4.19\frac{N_{Avag}\rho}{M}\alpha_e \qquad (4.1)$$

Here, n is the refractive index, α_m is the molecular polarizability, M is the molar mass, ρ is the density and N_{Avag} is the Avogadro constant. This equation assumes that the molar polarization is constant for a particular material irrespective of temperature and the pressure. Also it ignores the orientation of molecular dipoles in the applied field. These assumptions are justified at high frequencies (i.e. optical) as the orientation of molecules is too slow and hence the polarizability is purely electronic being insensitive to temperature and pressure even in highly condensed phases where intermolecular forces are large. For the equation to be valid any resonant absorption peaks need to be avoided as these cause the refractive index to be anomalously large. The Clausius-Mosotti equation was studied extensively by R.D. Shannon's research group in the early 1990's and so for greater understanding of its application then the reader is encouraged to read the references by Subramanian et al [45] and Shannon [46].

The polarizability (α_m) depends on the so-called structural units of which a material is composed [41]. In these calculations, it was assumed that the main structural unit for HfO_2 is HfO_8 [42].

The ALD hafnium oxide films deposited on a rapid thermal oxide IL were analysed by modelling the high-k as a Cauchy layer and by using the previously extracted thickness and optical constant values for the RTO interfacial layer. The thickness and refractive indices were extracted for all the samples using the procedure explained in chapter 2.10 for Cauchy layer fitting including an Urbach tail. The density for the ALD hafnium oxide films with rapid thermal oxide (RTO) interface was calculated using the polarizability for the HfO_8 structural unit calculated by Rignanese [41]. Rutherford back scattering (RBS) measurements were then performed on the samples to extract the hafnium concentration and the extracted density results were plotted against these and can be seen below in figure 4.14.

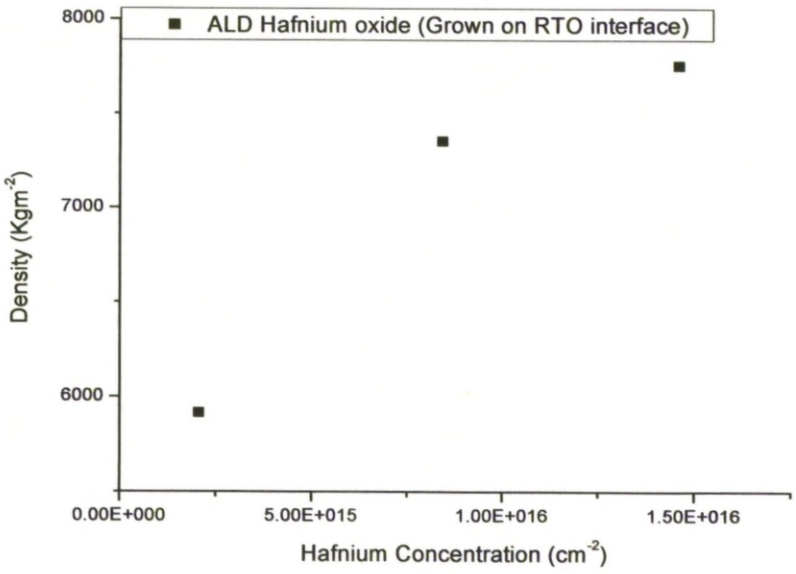

Figure 4.14 - Density variation for ALD hafnium oxide samples with different stoichiometry measured by Rutherford backscattering (RBS)

The density of pure bulk crystalline HfO_2 is 9680kg/m^3 [9, 10]. The density extracted for the ALD hafnium oxide films deposited on rapid thermal oxide varied between 5916-7753 kg/m^3 with the density increasing with the hafnium content. The likely reason for the wide range in calculated densities is due to

the lower stoichiometry and the lower physical density of the thinner hafnium oxide films. The stoichiometry is affected by the deposition method used to deposit a film and in the case of the ALD samples, is improved by increasing the number of deposition cycles (and hence increasing the film thickness).

4.6.2. Density extraction of hafnium silicates

The technique was next applied to HfSiO samples. The molecular mass (M) and polarizability (α_m) terms were obtained by considering specific values for a given composition. $\alpha_m = \alpha_m (M_{Ax}M'_{Bp}O_{Ay+Bq})$ for a mixed oxide (M_xO_y + M'_pO_q) using the additivity rule which states that

$$\alpha_m(M_{Ax}M'_{Bp}O_{Ay+Bq}) = A\alpha_m(M_xO_y) + B\alpha_m(M'_pO_q) \qquad (4.2)$$

This modification was required due to two reasons, the first was the presence of two forms of structural unit with different coordination within hafnium silicates, namely HfO_6 and HfO_8 and the second was that the silicon dioxide content in the layers needs to be considered. The prevalence of each of these structural units is expected to depend on the hafnium content, with low and high (>50%) hafnium content films being composed completely of HfO_6 and HfO_8 [42] respectively. The values of the density of the hafnium silicate films were calculated using polarizabilities from Rignanese [41], which varied between 7595 and 8319 Kg/m^3 for a variation of n between 1.65-1.95. The results presented in Fig. 4.15 show the density of the analyzed films if they were composed of a) only HfO_6 and b) only HfO_8; independent of composition in both cases. It can be seen that the densities increase with increasing hafnium content.

X-ray reflectometry (XRR) measurements were later carried out at IMEC on the hafnium oxide and hafnium silicate samples. A wavelength of 1.451Å was used and the incidence angle varied through the range 300-15000sec with a step-size of 50sec. The experimental data was modelled using a four-layer representation comprising of a silicon substrate, a silicon dioxide interfacial layer and two high-k layers representing a dense high-k layer and a less dense high-k overlayer. The properties extracted by XRR are shown in table 4.4.

Sample HfO2 ratio (x)	HfO2 /HfSiO2 layer thickness (nm)	HfO2 /HfSiO2 Dense layer density (g/cm3)	HfO2 /HfSiO2 overlayer Thickness (nm)	HfO2 /HfSiO2 overlayer density (g/cm3)	XRR averaged density (g/cm3)
1	3.40	9.87	0.198	8.56	9.80
0.7	2.75	7.52	0.425	6.66	7.41
0.5	2.78	6.57	1.575	1.81	4.86
0.3	2.28	5.28	1.227	0.85	3.73

Table 4.4 - Table showing the layer thicknesses and densities extracted by x-ray reflectometry for the two layer model and including the averaged layer density

The model of best fit for the experimental XRR results was obtained by considering the hafnium silicate as two layers, namely a dense high-k layer and a porous over-layer of density between 10-20% less than the dense layer. To allow comparison of the results from the XRR analysis and the single high-k layer density extracted using ellipsometry, the density measured using XRR was converted to a weighted average value using equation 4.8 below, where ρ_{OL} and ρ_{DL} are the densities of the dense layer and the over-layer and t_{OL} and t_{DL} are the thicknesses of the dense layer and over-layer respectively. Both the weighted average density value from equation 4.3 below and the density for the dense layer from the XRR results were plotted in figure 4.15 along with the density values for bulk HfO_2 and SiO_2 as a reference.

$$\frac{n^2-1}{n^2+2} = \left(\frac{4\pi}{3V_m}\right)\alpha_e = 4.19\left(\frac{\alpha_e}{V_m}\right) = 4.19\frac{N_{Avag}\rho}{M}\alpha_e \qquad (4.3)$$

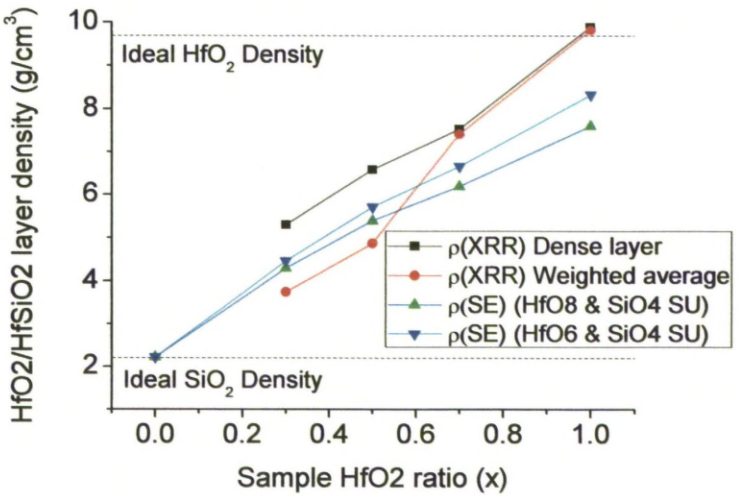

Figure 4.15 - Density variation for hafnium silicates thin films

It can be seen that the density extracted using SE was at most, 15% lower than that extracted using XRR. A possible source of error for the XRR results is the surface roughness of the films as the density values extracted for the porous over-layer for samples containing the highest proportion of SiO_2 dropped to below half that of bulk silicon dioxide. This suggests some limitations of the 4-layer model for the hafnium silicate layers at least for the case of hafnium oxide molecular fraction (x) less than 0.5.

Attempts were made to extract the properties of a thin porous over-layer added to the ellipsometry model, however as expected for such a thin film, further complexity added to the model resulted in large values of correlation between the high-k thickness, optical constants and the porous over-layer meaning that independent extraction was impossible. Therefore, it can be concluded that the one- layer high-k model for these samples represents the best that can be achieved with spectroscopic ellipsometry. It can be appreciated that further improvements to the accuracy of the density estimation technique will require better understanding of the chemistry occurring within the high-k dielectric materials. The accuracy of the density extraction technique using SE results could also be assessed by further study into the use of structural units within the films, given that structural units are used to describe a crystalline structure and

our films are thought to be amorphous. Densities of hafnium silicate films of various stoichiometry levels have been estimated previously [43] by combining film thickness data obtained by TEM, with the amount of hafnium deposited, as calculated from RBS measurements. A conclusion was drawn that the density of pure HfO_2 was lower than the ideal bulk value and also for thinner films [43]. However, the results reported here show densities for hafnium silicate films over a wide composition range. The results indicate that the density of MOCVD hafnium silicate films is inversely proportional to hafnium content. In comparison with the methodology presented in [43], this method has the advantage of being simpler, non-destructive and non-invasive and should be useful at least for comparative measurements in a production environment, if appropriately calibrated.

4.7. The static dielectric constant

The static dielectric constant has two contributions; one being the so-called electronic component, while the second is the lattice or phonon related component. Using the same terminology as in [41], the "zero" frequency static dielectric permittivity tensor $\varepsilon^{\infty}_{\alpha\beta}$ can be written as:

$$\varepsilon^{0}_{\alpha\beta} = \varepsilon^{\infty}_{\alpha\beta} + 4\pi \sum_{m} \frac{f^{2}_{m,\alpha\beta}}{\omega^{2}_{m}} \tag{4.4}$$

where $\varepsilon^{\infty}_{\alpha\beta}$ is the high frequency dielectric permittivity tensor and the second term accounts for the phonon contributions. From SE measurements between 240 – 1700 nm (1.57 – 0.17×10^{9} MHz), the high frequency part of the complex dielectric constant was extracted. The same samples were analysed using C-V measurements and a dielectric constant extracted from the accumulation capacitance. The phonon contribution can be then related to information obtained through atomic structure calculations – [44], and IR spectroscopy results available in the literature [42 and references therein].

The static dielectric constant obtained from C-V measurements and the high frequency dielectric permittivity from SE was substituted into equation 4.3 and the lattice contribution was extracted for all of the hafnium silicate samples. The results are presented in figure 4.16. The figure also contains data for the chemical oxide interfaced HfO_2 sample which has different stoichiometry, as

well as data for pure SiO_2. These results can give an indication of the contributions brought in by different atomic entities and the impact upon the static dielectric permittivity. It can be seen that for the sample around a hafnium oxide content of 50%, a spurious peak is apparent in the results.

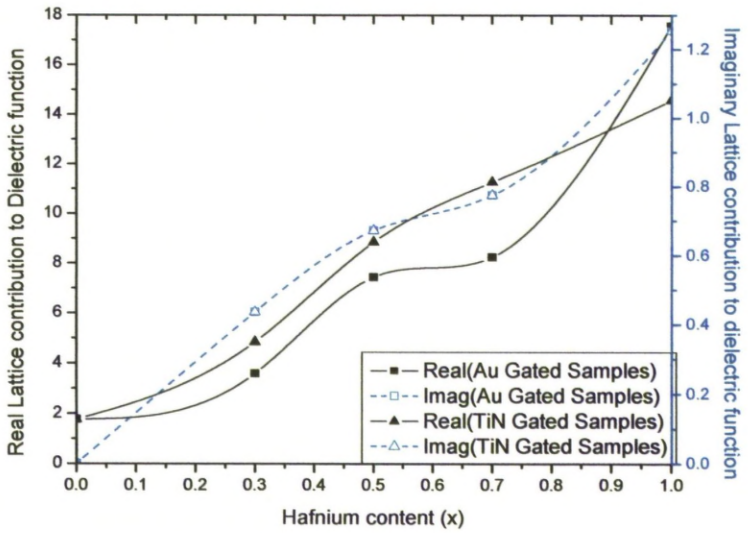

Figure 4.16 - Lattice contribution to the dielectric constant for hafnium silicate films of various compositions

MEIS was carried out on the samples at the Daresbury MEIS facility using a 100keV He ion beam. The double alignment procedure was performed to give true surface sensitivity by reducing scattering from atoms deep within the silicon substrate as explained in the MEIS section. The MEIS spectra for the hafnium silicate layers of all nominal compositions can be seen in figure 4.17. Here it can be seen that the hafnium related peak decreases in both intensity and broadening as the silicon dioxide content increases, proving that the hafnium silicate layers become, as expected, lower in hafnium content and the films are physically thinner, as seen SE analysis. A shoulder on the silicon peak is also apparent, increasing in magnitude as the silicon dioxide content of the layer is increased. The SIMNRA program, as mentioned in chapter 3.3, was used to model the energy lost by the ions thus allowing depth profiles to be extracted. The thicknesses of the layers can be seen in table 4.5 below. The hafnium oxide to silicon dioxide ratios for all of the samples were extracted from the silicon peak in the MEIS spectra and are also shown in table 4.5 below. It can be seen that the hafnium oxide content of the layers is accurate for high hafnium oxide

content but there is quite a considerable difference with the low hafnium oxide content. Most notable is the result for the nominally 50% hafnium oxide layer, which at 65% HfO_2 is very close to the content of the nominally 70% HfO_2 layer. This could explain the spurious peak in the extracted values for the lattice contributions to the relative permittivity seen in figure 4.16 as this would shift the nominal 50% HfO_2 point closer to the nominal 70% HfO_2 point giving the graph a non-linear but possibly more physically reasonable 'dispersion' shape

Nominal HfO_2:SiO_2 ratio	**Extracted HfO_2:SiO_2 ratio**	**Silicate thickness (nm)**	**Interfacial chemical oxide thickness (nm)**
100:0	100:0	4.2	1.0
70:30	71:29	3.3	1.5
50:50	65:35	3.1	1.6
30:70	35:65	2.5	1.8

Table 4.5 - Layer thicknesses and hafnium oxide to silicon dioxide ratio extracted from MEIS

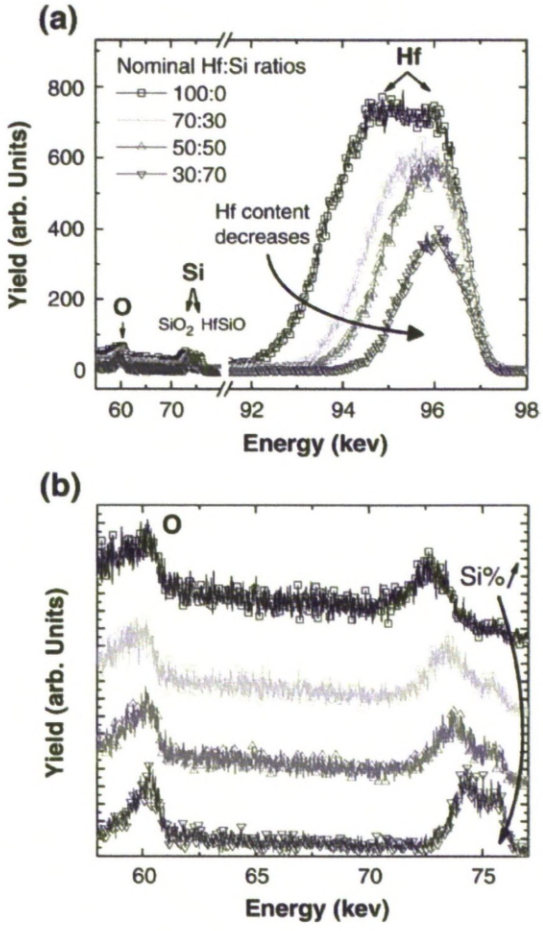

Figure 4.17 - Measured MEIS spectra from HfSiO layers on Si (100) with SiO2 interlayer (a) and zoom on lower energies where a second peak is seen when Si content in high-k layers increases (b)

Cross-sectional transmission electron microscopy (TEM) was carried out on all of the samples by collaborators at the University of Manchester. The samples were prepared using standard argon ion milling methods and TEM was performed using a Gatan double tilt liquid N_2 cooling rod in an FEI Tecnai F20 operated at 200kV. The TEM images for the 100% HfO_2 and the nominally 50% HfO_2 samples are shown below in figure 4.18. The hafnium based layer appears dark in the TEM images, however in some of the images; the upper surface of the layer appeared lighter than the bulk of the layer. A possible explanation for this effect is diffusion of glue into the hafnium based layer during specimen preparation, possibly exacerbated by surface roughness.

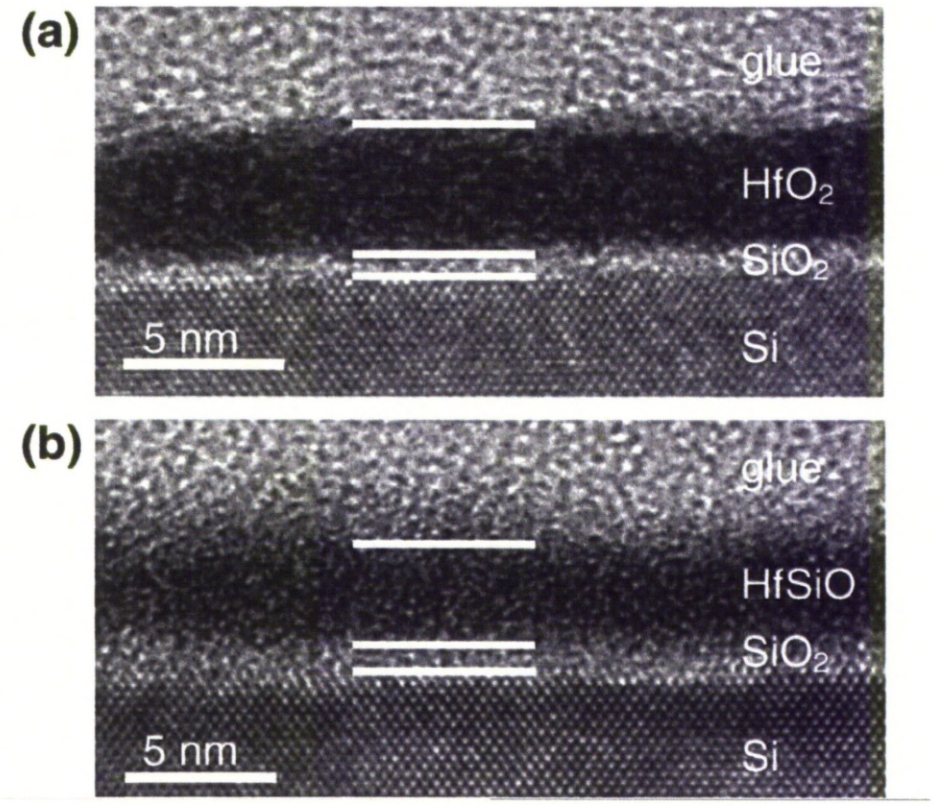

Figure 4.18 - Cross TEM images for the a) 100% hafnium oxide content layer and b) the 50% hafnium oxide content layer samples

Thickness results for all of the methods, SE, MEIS and TEM, were plotted on the same graph of film thickness versus hafnium oxide content and this can be seen in figure 4.19 below to allow comparison. It can be seen that the thicknesses extracted using the three different methods concur to within fractions of a nanometre. The bigger differences between the TEM thicknesses than the others, although extremely small, could be due to the fact that SE and MEIS measure average thickness and TEM allows analysis the thickness to be viewed directly by the user.

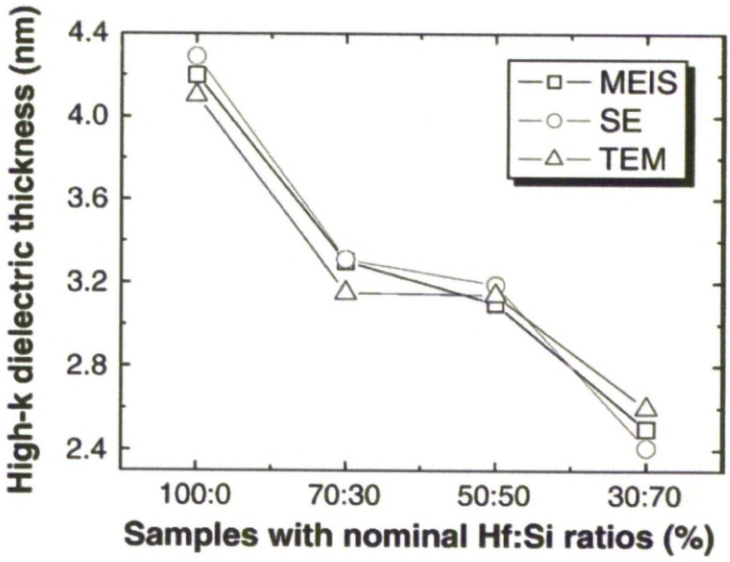

Figure 4.19 - Graph showing extracted high-k layer thickness versus the nominal Hf:Si content ratios for three different analysis techniques

4.8. Conclusions

In this chapter, analyses for hafnium oxide and hafnium silicate films deposited by ALD have been presented. As the silicon content of a hafnium silicate film is increased both the optical and electrical permittivity is observed to decrease. The band-gap values for the hafnium silicate films were seen to increase with silicon content but the leakage currents measured through the films were seen not to be solely dependent on composition. Defect energy levels were observed in the hafnium silicates which occurred deeper in the band-gap as the silicon content was increased, however, a silicon content of 30% was seen to contain a shallow trap possibly caused by the hafnium 'impurity' within the film. An estimation technique was used to assess the density decrease in a hafnium silicate as the silicon content was increased and the technique was verified to have an error of less than 15% compared to a standard density estimation method.

The lattice contribution of the hafnium silicate material to the static permittivity was extracted and an erroneous feature was proven by MEIS, to be caused by the variation of actual film composition from what was predicted due to deposition parameters. Finally SE, MEIS and TEM were compared as thickness measurement techniques and it was observed that they were accurate within

experimental error. There was a slight variation between TEM and the other techniques and this was predicted to be due to SE and MEIS measuring average thickness.

4.9. References

[1] K. Onishi, C. S. Kang, R. Choi, H-J. Cho, S. Gopalan, R. Nieh, S. Krishnan and J.C. Lee., VLSI Symp. Technical Digest, pp. 22-23, (2002)

[2] C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White, and P. Tobin, VLSI Symp. Technical Digest, pp. 9-10, (2003)

[3] Y. Kim, C. Lim, C. D. Young, K. Matthews, J. Barnett, B. Foran, A. Agarwal, G. A. Brown, G. Bersuker, P. Zeitzoff, M. Gardner, R. W. Murto, L. Larson, C. Metzner, S. Kher and H. R. Huff, VLSI Symp. Technical Digest, pp. 167-168, (2003)

[4] T. Kauerauf, R. Degraeve, E. Cartier, B. Govoreanu, P. Blomme, B. Kaczer, L. Pantisano, A. Kerber, and G. Groeseneken, IEDM Technical Digest, pp. 521-522, (2002)

[5] H. Kim, P.C. McIntyre and K.C. Saraswat, Appl. Phys. Lett., 82, pp. 106-108, (2003)

[6] G.D. Wilk, R.M. Wallace, Appl. Phys. Lett., vol. 74, pp. 2854-2856, (1999)

[7] R.M.C. de Almeida and I.J.R. Baumvol, Surf. Sci. Rep., vol. 49, pp. 1-114, (2003)

[8] M. Koyama, A. Kaneko, T. Ino, M. Koike, Y. Kamata, R. Iijima, Y. Kamimuta, A. Takashima, M. Suzuki, C. Hongo, S. Inumiya, M. Takayanagi and A. Nishiyama, IEDM Technical Digest, pp. 849-852, (2002)

[9] W. J. Qi, R. Nieh, B. H. Lee, K. Onishi, L. Kang, Y. Jeon, J. C. Lee, V. Kaushik, B. Y. Neuyen, L. Prabhu, K. Eisenbeiser, and J. Finder, VLSI Symp. Technical Digest, pp. 40-41, (2000)

[10] W. Zhu, T. P. Ma, T. Tamagawa, Y. Di, J. Kim, R. Carruthers, M. Gibson, and T. Furukawa, IEDM Technical Digest, pp. 463-466, (2001)

[11] G.D. Wilk, R.M. Wallace and J.M. Anthony, J. Appl. Phys., vol. 89, pp. 5243-5275, (2001)

[12] M. Ritala, K. Kukli, A. Rahtu, P. I. Raisanen, M. Leskela, T. Sajavaara, and J. Keinonen, Science, vol. 288, pp. 319-321, (2000)

[13] Z.M. Rittersma, E. Naburgh, T. Dao, A.H.C. Hendriks, W.E.A. Beling, E. Tois, E. Vainonen-Ahlgren, M. Tuominen and S. Haukka, Electrochem. Solid-State Lett., vol. 6, F21, (2003)

[14] M. Lemberger, A. Paskaleva, S. Zurcher, A. J. Bauer, L. Frey, H. Ryssel, et al., Microelectron. Rel., vol. 45, pp. 819-822, (2005)

[15] M. Balog, M. Schreiber, S. Patai and M. Michman, J. Cryst. Growth, vol. 17, 298-301, (1972)

[16] P. J. Harrop and D.S. Campbell, Thin Solid Films, vol. 2, pp. 273-292, (1968)

[17] J. Robertson, J. Vac. Sci. Tech. B, vol. 18, pp. 1785-1791, (2000)

[18] Z. Xu, M. Houssa, S. De Gendt, M. Heyns, Appl. Phys. Letts., vol. 80, no. 11, pp. 1975-1978, (2002)

[19] Buiu, O.B., Spectroellipsometric study of hafnium silicates, Private presentation

[20] H. Takeuchi and Tsu-Jae King, Mat. Res. Soc. Symp. Proc., Vol. 811, D7.6

[21] S. Dueñas, H. Castán, H. Carcía, J. Barbolla, K. Kukli, J. Aarik, M. Ritala, M. Leskelä, Microelectron. Rel., vol. 45, pp. 949-952, (2005)

[22] G.D. Wilk, R.M. Wallace and J.M. Anthony, J. Appl. Phys., vol. 87, pp. 484-492, (2000)

[23] G. Lucovsky, Y. Wu, H. Niimi, V. Misra and J.C. Phillips, Appl. Phys. Lett., vol. 74, pp. 2005-2007, (1999)

[24] G.D. Wilk, R.M. Wallace, J.M. Anthony, J. Appl. Phys., Vol.89, No.10, pp. 5243-527, (2001)

[25] S. Xing, N. Zhang, Z. Song, Q. Shen, C. Lin, Microelectronic Engineering, vol. 66, pp. 451–456, (2003)

[26] M.-H. Cho, K. B. Chung, C. N. Whang, D. W. Lee, and D.-H. Ko, Appl. Phys. Lett., vol. 87, 242906, (2005)

[27] S. Sayan, N. V. Nguyen, J. Ehrstein, J. J. Chambers, M. R. Visokay, M. A. Quevedo-Lopez, L. Colombo, D. Yoder, I. Levin, D. A. Fischer, M.

Paunescu, O. Celik, and E. Garfunkel, Appl. Phys. Lett., vol. 87, 212905 (2005)

[28] K. B. Chung, C. N. Whang, M. -H. Cho, C. J. Yim, and D. -H. Ko, Appl. Phys. Lett. vol. 88, 081903, (2006)

[29] Hag-Ju Cho, Hye Lan Lee, Hong Bae Park, Taek Soo Jeon, Seong Geon Park, Beom Jun Jin, Sang Bom Kang, Yu Gyun Shin, U-In Chung and Joo Tae Moon , J. J. Appl. Phys., vol. 44, No.4B, pp. 2230-2234, (2005)

[30] Intel Corp. website, http://www.intel.com/technology/architecture-silicon/45nm-core2/index.htm (Web address correct in Sept. 2009)

[31] K.J. Yang and C. Hu, IEEE Trans. Electron. Devices, vol. 46, pp. 1500–1501, (1999)

[32] S. Dueñas, H. Castán, H. García, J. Barbolla, K. Kukli, J. Aarik, M. Ritala, and M. Leskelä, Microelectron. Rel., vol. 45, pp. 949–952, (2005)

[33] H. Kato, T. Nango, T. Miyagawa, T. Katagiri, K. S. Seol, Y. Ohki, J. Appl. Phys., vol. 92, pp. 1106–1111, (2002)

[34] A. Callegari, E. Cartier, M. Gribelyuk, HF Okorn-Schmidt and T. Zabel, J. Appl. Phys., vol. 90, pp. 6466–6475, (2001)

[35] G. D. Wilk, R. M. Wallace, and J. M. Anthony, J. Appl. Phys., vol. 87, pp.484–492, (2000)

[36] Y. Harada, M. Niwa, L. Sungjoo, and K. Dim-Lee, Symp. VLSI Technical Digest, pp. 26–27, (2002)

[37] M. Lemberger, A. Paskaleva, S. Zurcher, A.J. Bauer, L. Frey and H. Ryssel, Microelectron. Rel., vol. 45, pp. 819–822, (2005)

[38] J. Kim and K. Yong, J. Cryst. Growth, vol. 263, pp. 442–446, (2004)

[39] K. B. Chung, C. N. Whang, M. -H. Cho, C. J. Yim, and D. -H. Ko, Appl. Phys. Lett. vol. 88, 081903, (2006)

[40] C. Krug and G. Lucovsky, J. Vac. Sci. Techn. B, vol. 22, pp. 1301–8, (2004)

[41] G-M. Rignanese, J. Phys.: Cond. Matter., vol. 17, R357-379, (2005)

[42] G. Lucovsky and G-B. Rayner Jr, Appl. Phys. Lett., vol. 77, pp. 2912-2914, (2000)

[43] R. L. Puurunen, A. Delabie, S.V. Elshocht, M. Caymax, M.L. Green, B. Brijs, O. Richard, H. Bender, T. Conrad, I. Hoflijk, W. Vandervorst, D.

Hellin, D. Vanhaeren, C. Zhao, S. De Gendt and M. Heyns, Appl. Phys. Lett., vol. 86, 073116, (2005)

[44] R.L. Puurunen, Chem. Vapor. Deposition, vol. 9, pp. 249-257, (2003)

[45] R.D. Shannon, M.A. Subramanian, T.H. Allik, H. Kimura, M.R. Kokta, M.H. Randles and G.R. Rossman, J. Appl. Phys., vol. 67, pp. 3798-3802, (1990)

[46] R.D. Shannon, J. Appl. Phys., vol. 75, pp. 348-366, (1993)

Chapter 5

Gadolinium oxide based gate dielectrics

5. Gadolinium oxide based gate dielectrics

5.1. Introduction

In this chapter, optimisation of the deposition and processing of dielectric stacks including gadolinium oxide, Gd_2O_3, is attempted.

The first section shows the development of a precision etching process for the controlled thinning of ultrathin silicon dioxide layers. These etched interfacial layers then are then used to analyse the effect of interfacial layer thickness on the CET of a stack containing e-beam evaporated gadolinium oxide after high temperature rapid thermal anneal (RTA).

The second section describes the realisation of dielectric stacks which are viable for the 45nm technology node, using atomic layer deposition (ALD) produced gadolinium oxide layers. The two methods used to produce the gadolinium silicate layers involve incorporation of Si during deposition and also as a result of high temperature anneals.

5.2. Literature review for gadolinium oxide and silicate

Gadolinium oxide, Gd_2O_3, is a promising replacement gate dielectric which has been extensively investigated [1, 2, 3-5]. Gadolinium oxide deposition has been demonstrated by MOCVD [4], anodic oxidation [6], thermal oxidation [7, 8], MBE [9, 10], e-beam evaporation [11, 12] and ALD [13-15]. The use of Gd_2O_3 dielectric layers has been reported for Si [2, 10] and also possible candidates for high mobility substrates including III-V compounds, such as GaN [16] and GaAs [17]. The effective dielectric constant of the Gd_2O_3 films is in the range of 7–23 and is strongly dependent on deposition method and process conditions [4]. The reported values for band gap also vary from 5.2 eV [18, 19], 5.3 eV [20] and up to 5.9 eV [21]. The conduction and valence band offsets to silicon are larger than 2 eV [9, 19]. It has been found that Gd_2O_3 decomposes according to the equation $2Gd_2O_3+3Si=4Gd+3SiO_2$ with $\Delta G>100$ kJ, and is expected to be thermodynamically stable on silicon up to 1000°C [22]. Another benefit for gadolinium over many of the rare earth oxides is that it is a single valence metal ion (+3) and hence forms only a single oxide (Gd_2O_3). The oxide also does not exhibit any intermediate metastable states whilst reacting with oxygen.

It is also possible to produce epitaxial layers of gadolinium oxide as the lattice parameter of Gd_2O_3 is 1.081nm in its bixbyite phase. Silicon has a lattice constant a_{Si} of 0.545 nm, then $2x(a_{Si})$ is only 0.4% larger than a_{Gd2O3}. Results have been published showing the electrical properties of epitaxial Gd_2O_3 thin films grown by MBE [9, 23], with EOT< 1 nm and leakage current density below 1 mA/cm^2. Referring to considerations set out in chapter 1, these figures exceed the requirements for low standby power applications which are predicted for the 32 nm node. Careful control of the thermodynamic parameters, such as oxygen chemical potential results in the interfacial layer change from oxide-like to a silicate-like, and thus leads to larger k and lower leakage [9]. The effect of rapid thermal annealing processes (RTA) on the structural and electrical properties of epitaxial Gd_2O_3 layers grown on Si has been discussed in [24]. The increase in CET for the layers due to interfacial layer growth has been shown to be significantly reduced by capping the high-k stack with amorphous-Si prior to RTA.

5.3. E-beam evaporated gadolinium oxide films

5.3.1 Introduction

The characterisation and analysis of gadolinium oxide films deposited using e-beam evaporation is detailed in this section. Furthermore, a process is developed to produce higher stability, silicate layers which could be formed within a CMOS thermal budget cycle. The idea was to optimise the interfacial layer, which would be mostly consumed as mixing with the hi-k layer took place during the thermal processing. The optimisation step involved discovering the optimum concentration of an aqueous hydrofluoric (HF) acid solution for etching thin 4nm SiO_2 films thermally grown on silicon so as to produce ultrathin high quality thermal oxides for use as interfacial layers. Having identified the optimum concentration, the time dependence of the etching process was established. Finally, gadolinium oxide layers were deposited on the samples using e-beam evaporation to study the effect of reducing the interfacial layer thickness on the physical and electrical properties of the gadolinium silicate film produced via high temperature annealing.

5.3.2. Experimental details

The etching experiments were carried out on 2x2cm silicon substrates, which had received a standard clean (SC1 and SC2 with HF dip) and were thermally oxidised with either a relatively thick nominally 100nm or a relatively thin, nominally 4nm silicon dioxide layer.

5.3.3. Wet etching 100nm thick silicon dioxide films

The first task was to conduct HF etching of 100nm thick SiO_2 films to find a solution with suitable etch rate and etch uniformity. Before and after etching, ellipsometry was carried out at a single wavelength of 632nm, at 9 positions on the 2x2cm sample squares. The film thicknesses were extracted assuming bulk-like properties for the silicon dioxide film.

The first etch on Sample 1, involved immersion of the wafer in a 1%HF solution for 5 minutes. Results are recorded in table 5.1 and illustrate the effect of stirring. It is apparent that 26% more oxide film is removed from Sample 1 (unstirred) compared to Sample 7 (stirred) with other conditions remaining the same. The unstirred film was also much less uniform, with 19% non-uniformity compared to less than 1% for stirred films. In all subsequent tests, the solution was stirred for 10s/min throughout the immersion.

To find the optimum concentration for etching thin SiO_2, the 100nm samples were immersed in various concentrations of HF solution for a set time of 5 minutes. The three chosen concentrations were 0.25, 0.5 and 1%HF and the oxide thicknesses can be seen before and after annealing in table 5.1 below.

Sample reference	Oxide thickness before etch (nm)	HF concentration (%)	HF immersion (etching) time (s)	Conditions	Oxide thickness after etch (nm)
Sample 1	99.6±0	1.0	300	Unstirred once immersed	57±11
Sample 4	100±0	0.25	300	Stirred before immersion and 10s every minute	93±0
Sample 6	101±1	0.5	300		89±0
Sample 7	100±0	1.0	300		77±0

Table 5.1 – Etching procedure details and film measurements for 100nm thick silicon dioxide samples

The extracted etch rates are shown in table 5.2 below and are also plotted in figure 5.1 as a function of HF concentration. It can be seen that the relationship between the etching rate and the HF concentration in the range under study is linear with a slope of 6.95×10^{-11} $ms^{-1}\%^{-1}$. Note that the plot does not go through the origin indicating that the true relationship with lower HF concentrations studied is likely not to be linear.

HF concentration (%)	Time (s)	Etched material (nm)	Etch rate (nm/min)
1%	300	22.5	4.5
0.5%	300	12	2.4
0.25%	300	6.8	1.36

Table 5.2 - Table showing etch rates for 100nm thick silicon dioxide films for various concentrations of HF solution after fixed immersion time

It can be seen from figure 5.1 that an HF solution concentration of 0.15% produces an etch rate of 1nm/min, which would offer good control for etching these ultra-thin films. However the HF available is prepared in a concentration

of 50% and so to produce a solution of 0.15% would require 3ml of 50%HF in 997ml of deionised water. As the lowest concentration that could be measured accurately with the equipment available was 5ml, then to accurately produce 0.15% would require two dilutions. This would not be time efficient and also prone to error in reproducibility. Therefore 0.25% HF was chosen as it had an etch rate of 1.36nm/min and could be produced with a single dilution.

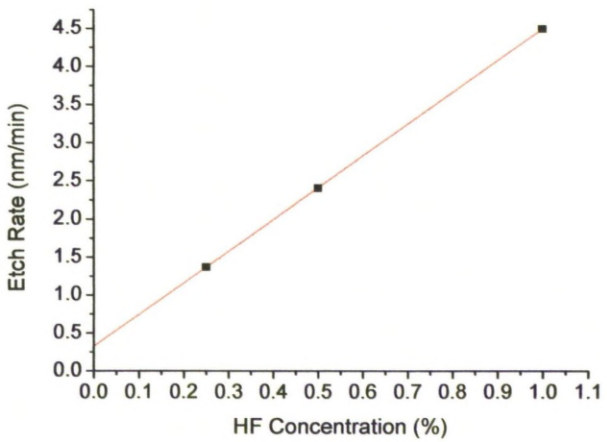

Figure 5.1 - Graph showing the relationship between etch rate and HF solution concentration for 100nm thick silicon dioxide films

5.3.4. Wet etching 4nm thick silicon dioxide films

The next task was to explore the etching of relatively thin 4nm silicon dioxide films to see if they behaved the same as the 100nm thick silicon dioxide films. For this stage it was decided that there would be three separate studies carried out with the eight, 4nm silicon dioxide films and these would be:

a) Three of the samples would be exposed to HF solutions with different concentrations for a fixed time to see the relationship between HF solution concentration and the etch rate for thin silicon dioxide films
b) Three of the samples would be exposed to the chosen 0.25%HF solution for varying immersion times to see the relationship between etching time and etched material for thin silicon dioxide films

c) Three of the samples would be exposed to the chosen 0.25%HF solution for the same immersion time to check for reproducibility.

The ellipsometry was carried out for a wavelength range between 240-1000nm for a fixed angle of 75° to obtain accurate thickness values. A J.A. Woollam M2000F instrument was used. For each sample, 5 measurements were taken, one at the centre of the sample and one at the centre of every edge. The samples were modelled as shown in figure 5.2. The first layer INTR_JAW, models a transition region with a higher silicon content than that of bulk silicon dioxide and hence a higher refractive index. The second layer from the substrate, SIO2_JAW, models bulk silicon dioxide.

Silicon dioxide	3nm
Si-SiO interface	1nm
Silicon substrate	Optically thick

Figure 5.2 - Spectroscopic ellipsometry model for a thermal oxide on silicon

When considering the experimental data for the samples before etching, the thicknesses for both the INTR_JAW and SIO2_JAW layers were selected as parameters to be fitted. For the samples after etching, the INTR_JAW thickness was fixed at the value before etching and the SIO2_JAW layer thickness was set as a fit parameter alone. The experimental details and the silicon dioxide thicknesses before and after HF etching are summarised in table 5.3 below.

Sample reference	Oxide thickness before etch (nm)	HF concentration (%)	HF immersion (etching) time (s)	Conditions	Oxide thickness after etch (nm)
S1 4nm	3.9±0.0	0.25	60	Solution was initially stirred before samples immersed. Samples where then stirred for 10s/min	2.6±0.0
S3 4nm	4.0±0.0	0.25	120		1.5±0.0
S4 4nm	4.0±0.0	0.25	180		0.9±0.0
S5 4nm	4.0±0.3	0.25	120		1.5±0.0
S6 4nm	4.0±0.0	0.25	120		1.5±0.1
S7 4nm	5.0±0.5	0.5	60		2.9±1.2
S8 4nm	4.2±0.3	0.15	60		3.5±0.2

Table 5.3 - Table showing the etch conditions and the initial and after immersion thicknesses for 4nm thick silicon dioxide films including uniformity

5.3.5.a. Effect of concentration on etch rate

Results for the effect of HF concentration were extracted from those in table 5.3 and are compiled in table 5.4 below. The samples were immersed in solutions of three different concentrations for the same fixed immersion time.

HF concentration (%)	Time (s)	Etched material (nm)	Etch rate (nm/min)
0.15	60	0.66	0.66
0.25	60	1.27	1.27
0.50	60	2.04	2.04

Table 5.4 - Table showing the results for varying HF solution concentration

The results for etch rate were plotted against the concentration of the etching solution for an immersion time of 60s and this is shown in figure 5.3 below. From figure 5.3, the relationship between etch rate and HF solution concentration is non-linear. The experiments were repeated and found to be reproducible. The etch rate for 0.25%HF solution is similar to that seen for

100nm silicon dioxide films, however the etch rate for 0.5%HF etching solution is much lower than for the 100nm films as shown in table 5.2.

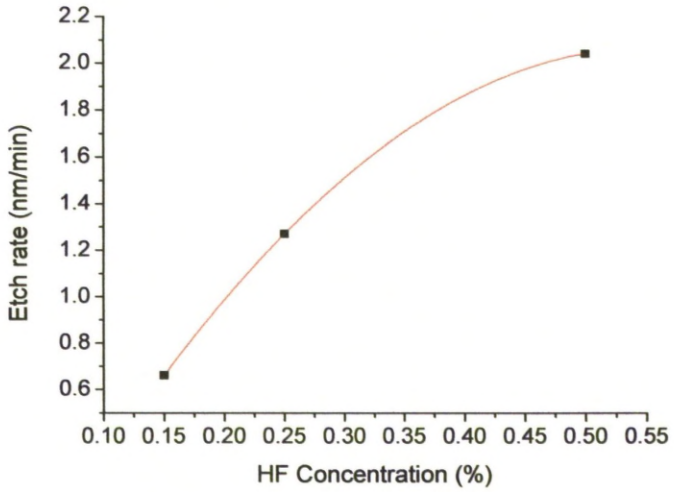

Figure 5.3 - Graph showing the relationship between the uniform etch rate and the concentration of the HF etching solution for an immersion time of 60s

5.3.5.b. Effect of immersion time

The results for immersion time were extracted from thos in table 5.3 and are compiled in table 5.5 below. The samples were immersed in the same concentration of HF solution for varying immersion times ranging between 1-3 minutes. The condition of 0.25%HF solution and 120 second etching time was repeated two further times to check for the repeatability of the etch, such as the consistency of the HF solution dilution and the experimental variables.

HF concentration (%)	Time (s)	Etched material (nm)	Average etch rate (nm/min)	Estimated etched material for last minute (nm)
0.25	60	1.27	1.27	1.27
0.25	120	2.50	1.25	1.23
0.25	120	2.50	1.25	1.23
0.25	120	2.45	1.23	1.18
0.25	180	3.07	1.03	0.59

Table 5.5 - Table showing the etching results for the samples studying immersion time alone

The etch rate data from table 5.6 was first plotted against immersion time to see if the etch rate was consistent with increasing etching time and this is shown in figure 5.4. It should be noted that etch rate is the average over the time the sample is immersed.

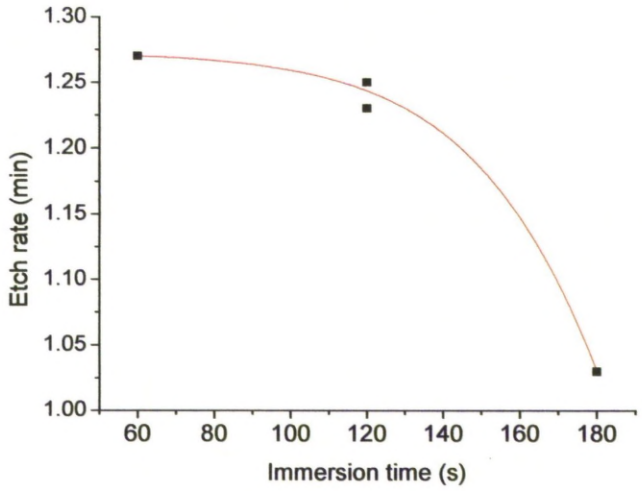

Figure 5.4 - Graph showing the average etch rate of a film with increasing etching time for a 4nm film in a 0.25%HF solution

For this reason it is better to view the results in terms of how much material is etched as a function of immersion time and this can be viewed in figure 5.5 below. From figure 5.5 it can be seen that the etch rate may be constant until an immersion time of 120 seconds, at which point it starts to decrease, although more data points are required to confirm this. To examine the effect further, film thickness after etching was plotted against the immersion time and is shown in figure 5.6. It can be seen that the etch rate decreases for a film thickness of about 1.0-1.5nm. This effect could relate to a region of slower HF rate, possibly associated with related to a higher density region, or it could relate to a significant percentage of the HF in the solution being consumed.

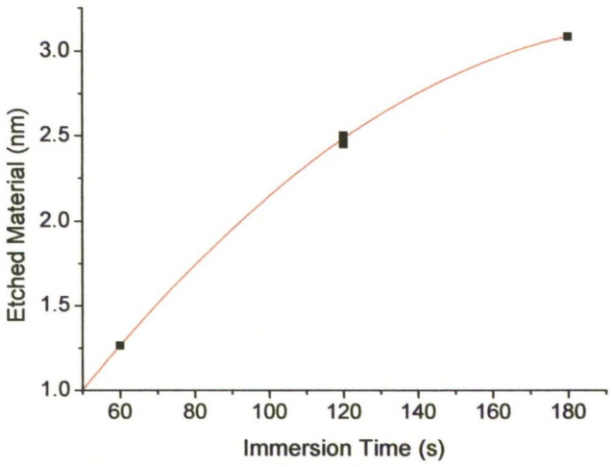

Figure 5.5 - Graph showing the total etched material over the sample immersion time for a 0.25%HF solution

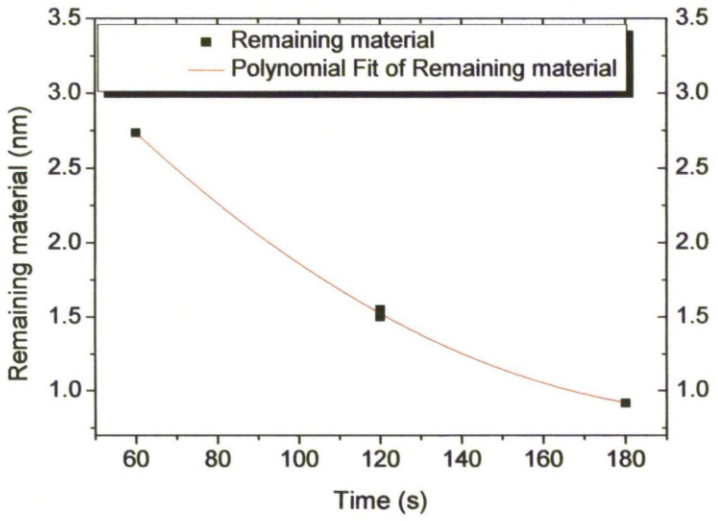

Figure 5.6 - Graph showing the remaining silicon dioxide on the sample as a function of the immersion time for a 4nm silicon dioxide layer in 0.25%HF solution

5.3.6. Experimental samples for Gd_2O_3 deposition

The motivation for the etching study was to investigate the effect of varying the thickness of the interfacial thermal oxide on the physical and electrical properties of annealed high-k MOSCAPs with Gd_2O_3 high-k layers. Therefore a new set of standard cleaned (SC1, SC2 and HF dipped) silicon samples were thermally oxidised to produce a 4nm silicon

dioxide layer. The samples were then etched as described in the table below with two samples being prepared of each IL thickness:

Interfacial layer thickness (nm)	Immersion time in 0.25%HF (s)
4nm	None
3nm	50
2nm	91
1nm	165

Table 5.6 - Immersion time in 0.25%HF to produce required IL thickness from 4nm silicon dioxide

The samples were measured before and after etching with the same ellipsometry procedure as for the 4nm thermal oxides in section 5.2.4-5.2.5 and the results are shown in table 5.7 overleaf.

Sample name	Thickness before etch (nm)	HF concentration (%)	Immersion time (s)	Conditions	Thickness after etch (nm)
S9_4nm/ PN8_#1	3.82	0.25	Unetched	Solution was initially stirred and was stirred with sample 10s for every minute during immersion	Unetched
S10_4nm/ PN8_#2	3.81	0.25	Unetched		Unetched
S11_4nm/ PN8_#3	3.77	0.25	(1)		1.76±0.2
S12_4nm/ PN8_#4	3.80	0.25	(1)		1.98±0.1
S13_4nm/ PN8_#5	3.82	0.25	50		2.63±0.03
S14_4nm/ PN8_#6	3.82	0.25	50		2.67±0.09
S15_4nm/ PN8_#7	3.82	0.25	165		1.04±0.09
S16_4nm/ PN8_#8	3.82	0.25	165		1.19±0.04

Table 5.7 - Table showing the results of the etch on the 4nm thermal oxide experimental samples showing measured thickness and uniformity

[1]*The solution for samples 11 and 12 was not mixed before sample immersion and so the etch rate was much higher. Samples were immersed for 50 s (expected etch = 1nm), however solution removed 2nm.*

From table 5.7, the etched material and average etch rates were extracted and these are shown in table 5.8 below.

HF concentration (%)	Time (s)	Etched material (nm)	Etch rate (nm/min)
0.25	50	1.19	1.428
0.25	50	1.15	1.380
0.25	165	2.78	1.011
0.25	165	2.63	0.965

Table 5.8 - Table showing the extracted results from the 4nm experimental samples immersed in 0.25%HF solution

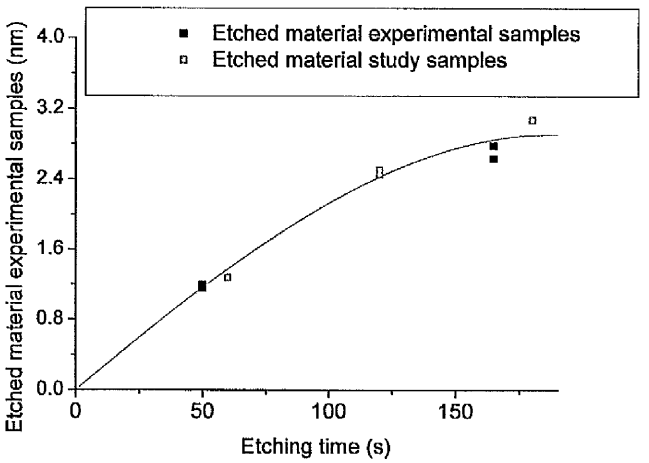

Figure 5.7 - Plot showing total etched material from a 4nm silicon dioxide film immersed in 0.25%HF solution for varying immersion times

The values for etched material with increasing immersion time were added to the results in figure 5.5 to improve the estimation of future etch steps and this can be viewed in figure 5.7. It can be seen from the plot that the relationship between etched material and immersion time is non-linear, suggesting that the silicon dioxide layer close to the silicon/silicon dioxide interface may etch at a slower rate, possibly due to increasing silicon content. There are also errors introduced when fitting both the data values as the experimental samples s9-s16

were slightly thinner (0.2nm) as-grown than the s1-s8 samples used for the etching study. Thus it might be expected that the initially thicker samples would encounter this slower etching region up to 10 seconds later than the thinner samples. The remaining thermal oxide for an initially 4nm layer after a given immersion time in 0.25% HF solution can be seen in figure 5.8 below

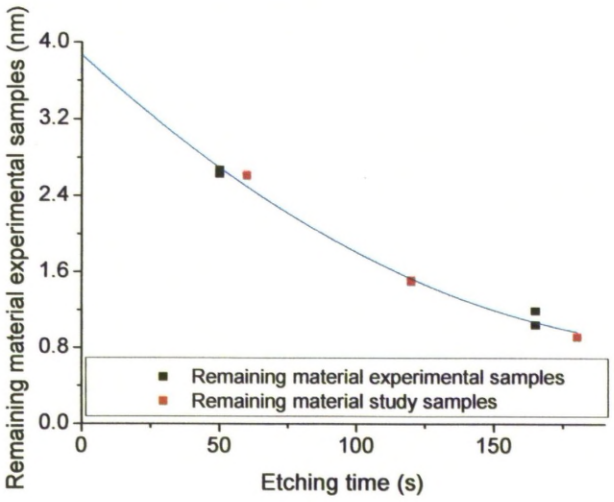

Figure 5.8 - Plot showing remaining material from a 4nm silicon dioxide film immersed in 0.25%HF for varying immersion times

After etching, a nominal 3nm Gd_2O_3 layer was deposited by e-beam evaporation. The stack was measured by spectroscopic ellipsometry using a fixed angle J.A. Woollam M2000F instrument. The experimental data was modelled as shown in figure 5.9. The existing model before deposition was used for the SiO_2 interfacial layer that is, using the model shown in figure 5.2 and table 5.7 for the interfacial layer thicknesses. The deposited high-k film was modelled as a Cauchy layer. The interfacial layer thicknesses were fixed. The high-k layer thickness for each of the samples was extracted and these we found to be very close to 3nm and uniform in all of the samples, as can be seen in table 5.9.

Cauchy layer	3nm
Silicon dioxide	2nm
Si-SiO interface	1nm
Silicon substrate	Optically thick

Figure 5.9 - Ellipsometric model for high-k dielectric on silicon with an interfacial layer of thermal oxide

Sample reference	Gadolinium oxide thickness (nm)
PN8_#2	3.30±0.03
PN8_#4	3.03±0.06
PN8_#6	3.00±0.06
PN8_#8	3.11±0.03

Table 5.9 - Table of gadolinium oxide layer thicknesses for all the samples

After SE measurement, the samples were split into two groups. The as-deposited samples, PN8_#1, 3, 5, 7 were used as controls and the samples PN8_#2, 4, 6 and 8 were divided in half for post-deposition annealing; half of each sample being annealed to 900°C and the other half up to 1000°C both in pure nitrogen for 1s. A 50 nm titanium nitride gate electrode layer was then deposited by sputtering and this was followed by lithography to form circular electrodes of diameter 100, 200, 300 and 400μm. The details for all of the samples are shown below in table 5.10.

Deposition Method	E-beam evaporation											
Initial SiO_2 thickness	Thermal Oxide (3.8 nm)			Thermal Oxide (2.5 nm)			Thermal Oxide (2 nm)			Thermal Oxide (1 nm)		
Gd_2O_3 thickness	**3 nm**											
Gate electrode (after RTA)	**TiN (50 nm)**											
RTA temperature ($1s/N_2$)	w/o	900°C	1000°C	w/o	900°C	1000°C	w/o	900°C	1000°C	w/o	900°C	1000°C
Sample number	#1	#2a	#2b (Damaged)	#3	#4a	#4b	#5	#6a	#6b	#7	#8a	#8b

Table 5.10 - Showing the interfacial layer thicknesses and post deposition anneal details of all the samples

Capacitance-voltage (C-V) measurements were carried out on all of the samples at four frequencies between 1 kHz and 100 kHz. The 1 kHz C-V measurements for the as-deposited samples with various thickness of interfacial layer are shown below in figure 5.10. It can be seen that the thickness of the interfacial layer on the as-deposited samples, has a large effect on the capacitance equivalent thickness (CET) of the gate stack, reducing the capacitance with increasing interfacial layer thickness as expected. The hump that can be seen in all of the traces is frequency dependent and is expected to be caused by P_{bo} centres associated with dangling bonds at the silicon surface.

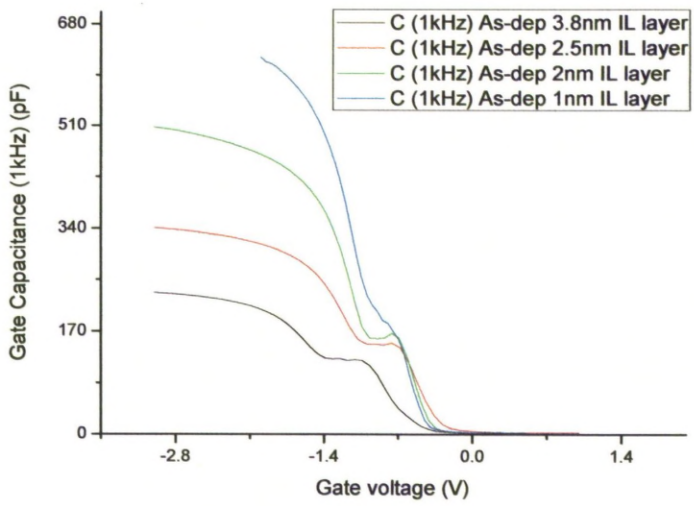

Figure 5.10 - Graph showing the 1 kHz CV measurements for identical as-deposited gadolinium oxide stacks with varying interfacial layer thicknesses

The CET of the stacks for all of the samples was calculated using the equation below where C_{acc} is the accumulation capacitance

$$CET = \frac{\varepsilon_0 \varepsilon_{SiO2} A}{C_{acc}} \tag{5.1}$$

The calculated CET values are plotted as in figure 5.11. It is clear from figure 5.11, that reducing the interfacial layer thickness causes reduction of the CET for the as-deposited samples. The annealing of thick, 10nm gadolinium oxide layers grown on a thick 4nm thermal oxide in nitrogen at 900°C for 1s, has previously shown a significant reduction in CET as the silicon dioxide layer is consumed, partially converting the high-k layer into a gadolinium silicate layer. This effect is apparent in figure 5.11 for the sample with the 4nm thermal oxide interfacial layer, where the 900°C RTA in N_2 for 1s, reduces the CET by 39% compared to the as-grown sample. However the samples with thinner interfacial layers reach a limiting minimum CET value of 2.6nm. Further reduction of the interfacial layer does not reduce the CET of the stack after annealing, by comparison to the as-deposited state, but instead increases it significantly. This is probably due to the presence of residual oxygen in the chamber during the RTA which causes the growth of interfacial SiO_x. As the interfacial oxide is consumed by the high-k layer at this annealing temperature, the SiO_x growth

compensates its loss. For initial interfacial oxides below 2.2nm (from as-deposited/PDA 900°C intercept on figure 5.11), the SiO_x growth actually causes the interfacial layer thickness to increase and hence suppresses the reduction in CET achieved by scaling the deliberately grown interfacial layer. This explanation is substantiated by the results from experiments at the higher annealing temperature of 1000°C. The limiting value of CET is much higher than that at the lower 900°C anneal, with a value of ~3.5nm. This is also likely to be due to residual oxygen in the annealing chamber but at 1000°C, the oxidation of the silicon surface is much more rapid than that at 900°C. The reduction in CET after annealing at 1000°C compared to the as-deposited film from figure 5.11, is limited to stacks with an initial interfacial thermal oxide layer of ~2.7nm. The thicknesses of the gadolinium oxide and silicate layers were extracted using HRTEM at the University of Liverpool and these values were used to calculate the k-values for the gadolinium silicate films. It was found that the as-deposited film had a k-value of ~14, which was improved during annealing to give increased k-values of between 15-17. The k-values measured after annealing correspond well to published figures which were reported by Gupta et al [36] who reported that the relative permittivity of gadolinium silicate to be ~15.8.

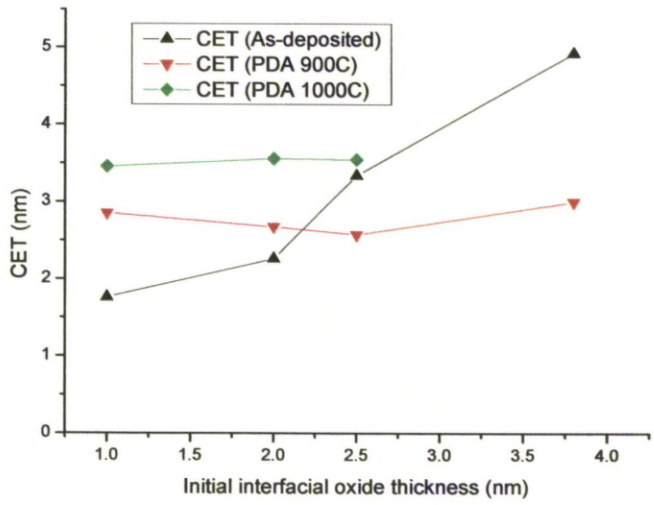

Figure 5.11 - Graph showing the CET versus the initial interfacial oxide thickness for different annealing temperatures

The current-voltage characteristics for all samples were measured. Current density at an effective bias of -1.5V was then extracted. The effective bias voltage accounts for the flat-band voltage of the stack. The current densities for the as-deposited samples are shown in figure 5.12 below, as a function of the initial interfacial silicon dioxide thickness. It can be seen that for every nanometre reduction in the interfacial layer thickness then the leakage current density increases by over an order of magnitude. After annealing for 1s in a nitrogen environment at 900°C, it can be seen in figure 5.11, that the CET of the stack does not change with the thinning of the initial interfacial layer below a value of 2.2nm. However, the leakage current density increases beyond this value. Recalling that the explanation for the limiting CET value was the occurrence of interfacial layer growth, then the increase in leakage current density could also be related to the interfacial layer. It is possible that the layer is defective due to its growth under the gadolinium oxide film with limited quantities of oxygen available in the annealing ambient. A similar trend is seen in the gadolinium oxide stacks that were annealed at the higher temperature of 1000°C in a nitrogen environment for 1s and the increase of leakage current density versus initial interfacial layer thickness can be seen for these samples in figure 5.14.

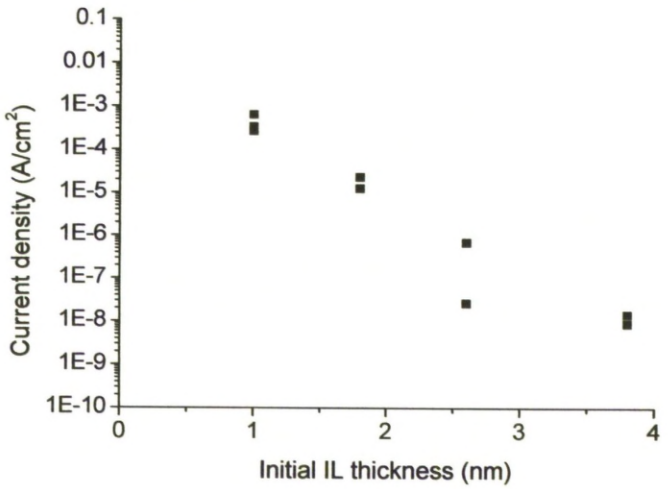

Figure 5.12 - Graph showing the leakage current density plotted versus the initial silicon dioxide IL thickness

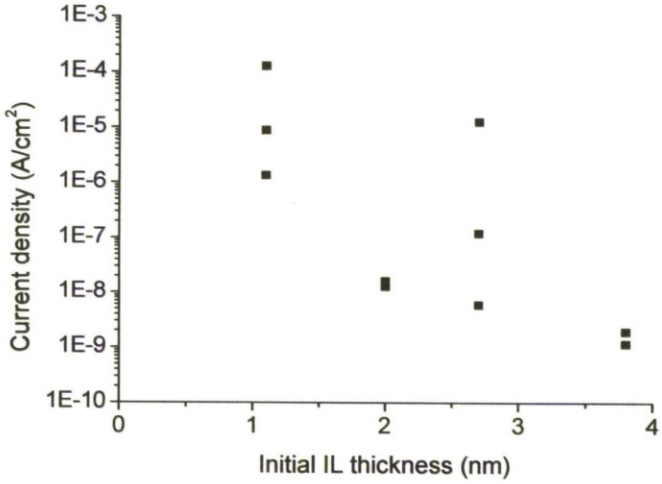

Figure 5.13 - Graph showing the leakage current density after annealing in N_2 for 1s at 900°C plotted versus the initial interfacial layer thickness

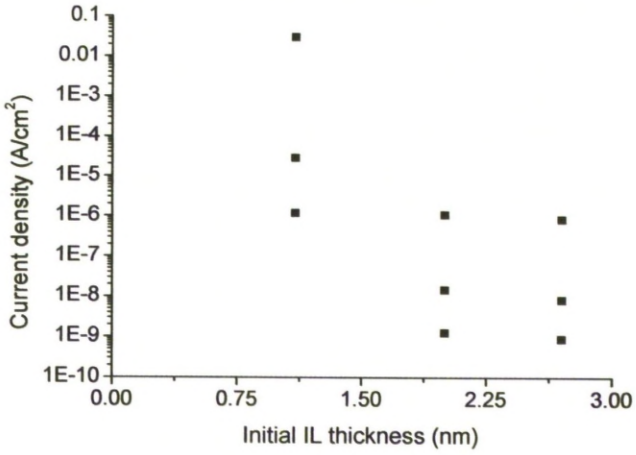

Figure 5.14 - Graph showing the leakage current density for the gadolinium oxide samples after annealing for 1s in N2 at 1000°C plotted versus the initial silicon dioxide IL thickness

All of the leakage current density values measured on the samples both as-deposited and annealed are plotted versus the extracted CET in figure 5.15 below. As the CET values for the annealed samples did not change significantly after the 900°C and the 1000°C RTA processes, then the leakage current densities for these sample sets are seen grouped on the plot. This graph allows

comparison of the leakage current densities after annealing, to those of the as-deposited stacks on equal terms by plotting the leakage against the CET for the stack. It can be seen that the lowest leakage comes from the annealed samples; even lower than the as-deposited sample with 4nm interfacial thermal oxide. However these samples do not display such low CET values as those as-deposited with the thinner interfacial oxides, which show considerably worse leakage. From these results, it can be concluded that to produce a gadolinium oxide stack with low CET but also low leakage, then the interfacial layer re-growth, which limits the CET during annealing, has to be reduced.

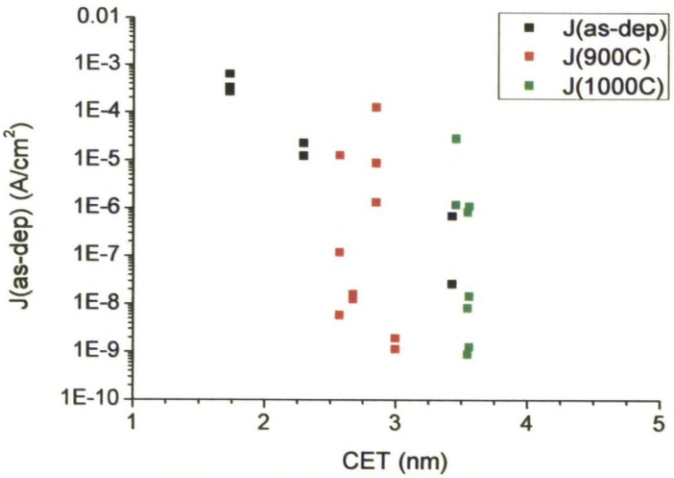

Figure 5.15 - Graph showing the leakage current density for the gadolinium oxide samples deposited on etched thermal SiO_2 IL before and after annealing at 900°C and 1000°C in N_2 for 1s plotted versus the CET for the samples.

5.3.7. Reduction of interfacial layer re-growth

A second set of samples was produced using the same etch and growth processes as described earlier in this chapter. The samples were etched to produce either a nominally 1nm or 3nm thermal oxide and then once again a 3nm gadolinium oxide was deposited using e-beam evaporation. In an attempt to reduce the interfacial oxide re-growth during the RTA step, the vacuum valve on the rapid thermal processor was left open so that the chamber was at a much lower pressure than the usual 0.5-1bar. All of the samples were exposed to a 900°C rapid thermal anneal in nitrogen ambient with the vacuum valve open, for one second (RTA2). One of each of the thicknesses was processed prior to the

RTA by annealing in a forming gas ambient at the relatively low temperature of 400°C for 30 minutes (RTA2+PDA) to passivate the silicon-silicon dioxide interface for good electrical performance. MOS capacitors where then formed by reactive sputtering of TiN from a titanium target followed by lithography to form gate electrodes. The gadolinium oxide samples analysed in section 5.3.6 had been exposed to a rapid thermal anneal at 900°C in a nitrogen environment for 1s but with the vacuum valve closed. They were analysed using TEM and EOT values were extracted and plotted in figure 5.16 labelled as RTA1 for comparison to the RTA2 and RTA2+PDA samples.

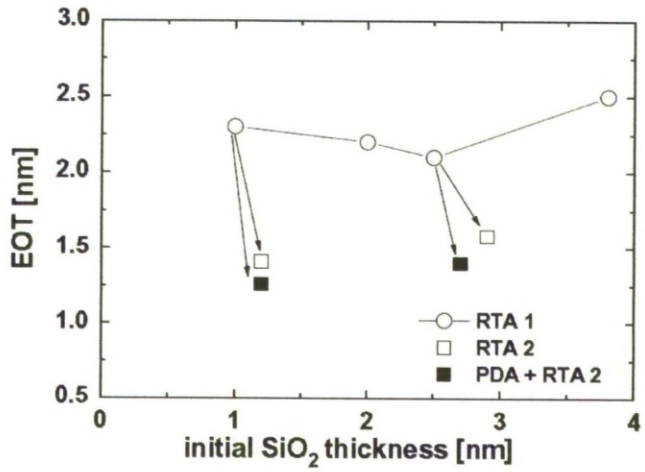

Figure 5.16 - Graph showing effect of varying the initial thermal oxide IL thickness and annealing procedure on the EOT for a stack containing a 3nm Gd_2O_3 layer

It can be seen in figure 5.16 that the revised RTA process with the vacuum valve open (RTA2) produces a film with a much lower EOT after annealing than with the vacuum valve closed (RTA1). Leaving the vacuum valve open during the RTA process must therefore minimise the effect of residual oxygen in the annealing chamber. The exact reason for this is unknown but could possibly be due to the nitrogen gas flow purging residual oxygen through the open vacuum valve. The samples which had been exposed to low temperature post-deposition anneals prior to the RTA process (PDA+RTA2) can also be seen to have a slightly lower EOT value than the samples that had not undergone this low temperature anneal. This is believed to be due to densification of the high-k layer during the low temperature anneal which slightly improves the CET after the RTA process.

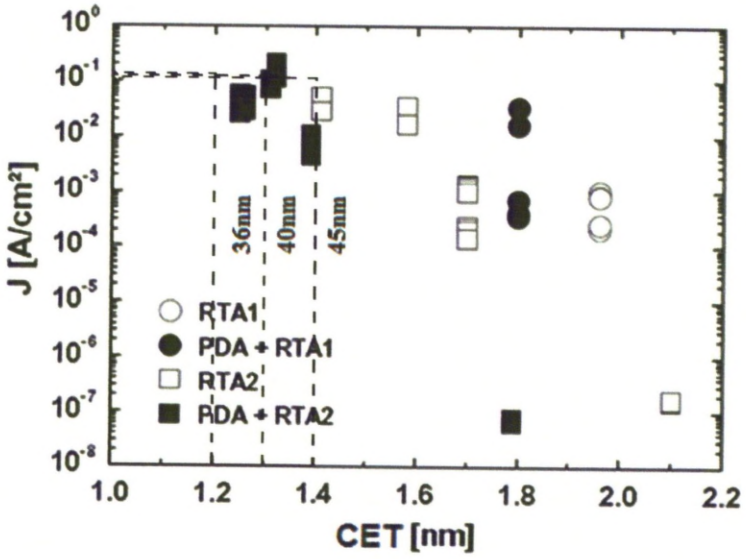

Figure 5.17 - Graph showing extracted current density values at an effective bias voltage of -1.5V plotted versus the CET of the stack for the samples exposed to different annealing processes. Dotted lines indicate requirements for specific ITRS technology nodes.

Current-voltage measurements were taken for all of the samples on multiple capacitors. The current densities at an effective bias voltage of -1.5V were plotted versus the CET of the stack and are shown in figure 5.17. For RTA1 combined with the PDA process, there is a limit to the CET of 1.8nm and these layers had leakage currents in the range 10^{-4}-10^{-3} A/cm^2. Much lower leakage current densities of less than 10^{-7}A/cm^2 are seen for the samples annealed in the RTA2 only process for similar CET values. This allows the CET to be reduced further to trade leakage current for an effectively thinner dielectric film. The best leakage current density compared to CET was measured on a sample exposed to both the RTA2 and the PDA treatments where a leakage current density of 0.02A/cm^2 was achieved for a CET of 1.3nm. The target for the Academic Cluster work package 2.3 in the European project PULLNANO was to produce a dielectric stack with a leakage current density of 0.01A/cm^2 for a CET of less than 1.5nm and it can be seen from figure 5.17 that this was achieved using a gadolinium oxide stack with these annealing processes.

5.4. Gadolinium based samples deposited using atomic layer deposition

5.4.1. Gadolinium silicate by atomic layer deposition

Gadolinium oxide (Gd_2O_3) films were deposited on n-type Si(100) wafers by atomic layer deposition (ALD), from $Gd[N(SiMe_3)_2)]_3$ precursor and H_2O. Carbon contamination during the deposition was assessed by Auger electron spectroscopy (AES) and found to be negligible. The wafer temperature was varied between 175-275°C, the number of cycles was between 75 and 300 and in all cases the precursor was heated to 175°C. The details of the samples are shown in the table 5.11.

Spectroscopic ellipsometry (SE) measurements were performed using a M2000UI VASE™ variable angle spectroscopic ellipsometer over a spectral range for multiple angles of incidence (65°-80°, 5° steps). The model used in the analysis of the experimental SE data is shown in figure 5.18 and consisted of a silicon substrate with a fixed 1nm SiO_2 layer to account for the native oxide interfacial layer on the samples and a Cauchy layer, with an Urbach tail to model the deposited gadolinium oxide layer.

Cauchy Layer	High-k film
SiO_2	IL
Silicon	Substrate

Figure 5.18 - Model used to represent the ALD gadolinium oxide samples in the WVASE32 software

Sample	Temp (°C)	Number of ALD cycles	Thickness (nm) (by weight gain)	Thickness (nm) (by ellipsometry)
A046	200	300	16	24.69
A047	200	300	11	19.14
A048	200	100	4	7.37
A049	200	100	3	4.255
A050	225	100	0	1.52
A051	225	100	2	4.42
A052	175	100	4	4.955
A053	175	100	10	17.65
A054	200	75	1	3.72
A055	200	75	3	3.57
A056	225	100	0	4.75
A057	225	100	2	4.62
A058	200	100	4	5.45
A059	200	100	5	5.20
A060	250	100	2	4.01
A061	250	100	3	4.12
A062	275	100	0	3.65
A063	275	100	0	3.67

Table 5.11- Table showing the simplified deposition parameters of the gadolinium oxide films including the film thickness extracted using spectroscopic ellipsometry

The thickness of the Cauchy layer was determined by limiting the analysed wavelength range to the transparent region in the near infra-red and fitting the thickness along with the first Cauchy parameter, as described in section 2.10. The extracted thicknesses are shown in table 5.11 above. The optical constants were then extracted by fixing the Cauchy layer thickness and fitting the

remaining Cauchy parameters. Finally the Urbach tail function was included in the fit to account for absorptions close to and including the band edge.

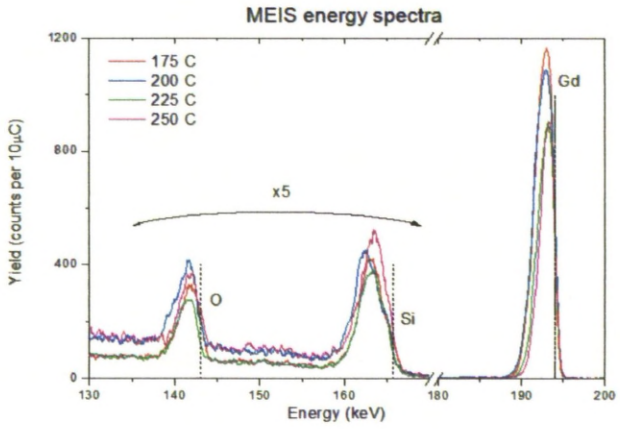

Figure 5.19 - MEIS energy spectra from along the [111] direction for samples grown at 175, 200, 225 and 250°C.

Medium energy ion scattering (MEIS) experiments [26] were carried out at the CCLRC Daresbury facility, using a nominally 200 keV He^+ ion beam and an ion dose corresponding to a charge of 10 μC. The samples were aligned to the ion beam along the [$\bar{1}\bar{1}$1] channelling direction and the electrostatic energy analyser was positioned to simultaneously record data from along the [111] and [332] blocking directions. These give scattering angles of 70.5° and 60.5° respectively and result in sub-nanometre depth resolution.

The MEIS spectra were used to produce compositional depth profiles. These must be considered as estimated profiles as it is assumed that the film layers have ideal density which for non-crystalline and non ideal samples is rarely the case; an ideal density for Gd_2O_3 of 7.07 g/cm^3 [27] was assumed. Energy scales have been converted to depth scales using established energy dependent inelastic stopping power values obtained from the SIMNRA program [28]. The results from MEIS showed a general trend for the growth of marginally thinner films for higher growth temperatures, which is confirmed from the measurements using spectroscopic ellipsometry.

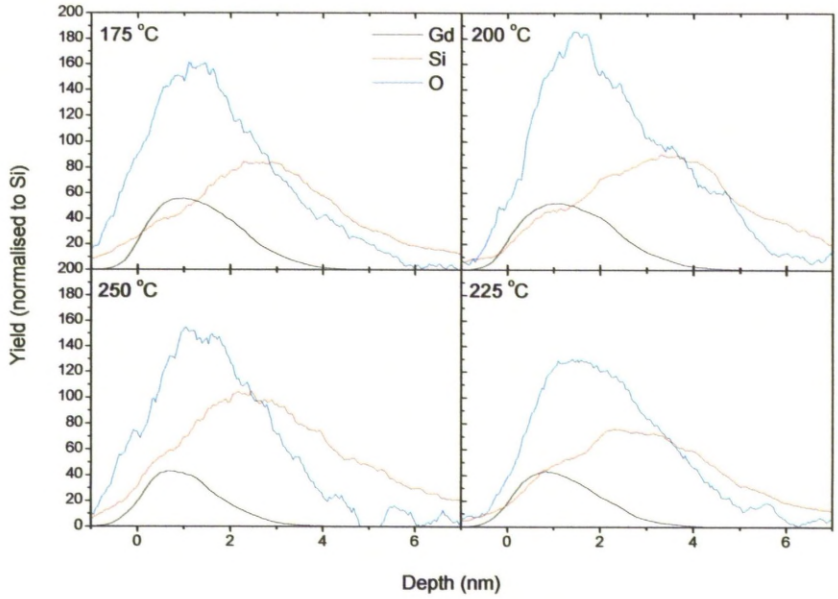

Figure 5.20 - Plots showing the depth profiles of the gadolinium oxide films extracted from the MEIS data for different wafer deposition temperatures

A shoulder on the high energy edge of the Si peak in the MEIS data indicated the presence of Si within the gadolinium oxide film as is expected when using this precursor. In fact, the Si is introduced deliberately to improve the thermal stability of the film. The silicon content was extracted using the technique outlined in chapter 3.3. The extracted silicon content was then plotted versus the deposition temperature and this is shown in figure 5.21 below. It can be seen that the relationship is not linear with the silicon content but increases steadily with increasing wafer deposition temperature.

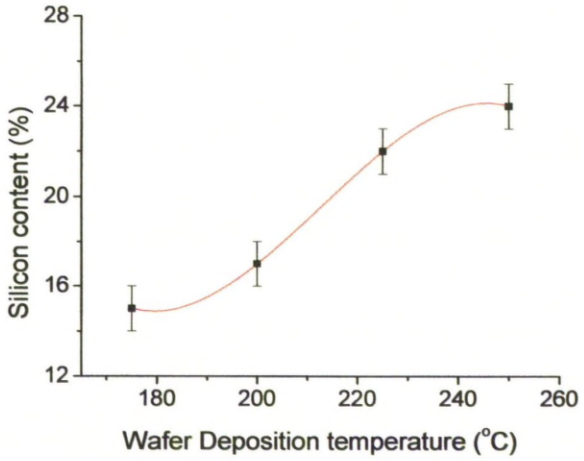

Figure 5.21 - Graph showing silicon content extracted from MEIS data plotted against deposition temperature for a fixed number of ALD cycles (100 cycles)

The thickness measured using spectroscopic ellipsometry was plotted versus the number of ALD cycles at the fixed deposition temperature of 200°C. A good linear dependence was obtained between the number of cycles and the thickness of the films, although the process conditions were seen to require further optimisation as it can be seen in figure 5.22 that a range of thickness values were obtained for samples deposited with the same number of ALD cycles.

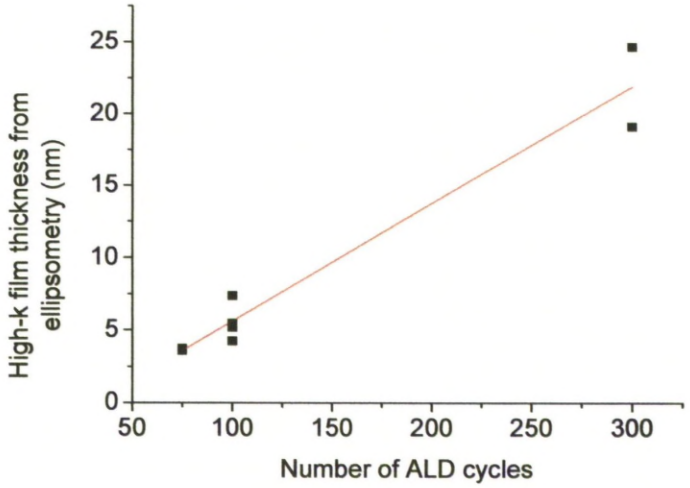

Figure 5.22 - Film thickness plotted against number of deposition cycles at 200°C.

The extracted values of the real part of permittivity were plotted versus the incident photon energy and these are shown in figure 5.23 below for the samples deposited with 100 ALD cycles. The highest values of ε_1 for these samples (2.7–3.27 over the spectral region analyzed) were obtained for the samples produced at 175°C.

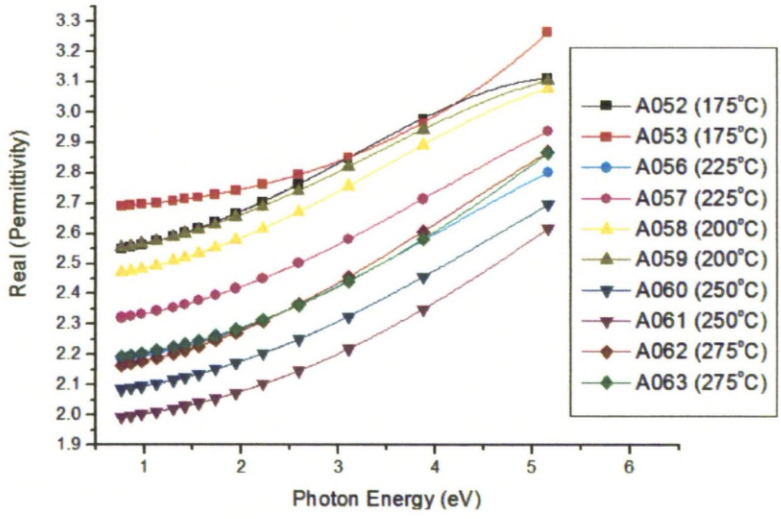

Figure 5.23 - Plot showing the extracted real permittivity functions for the 100 cycle films plotted against incident photon energy

The extracted value for the real part of permittivity at the longest wavelength was then taken for each of the samples, to assess the relationship between the deposition parameters and the density. Using the electronic polarizability value taken from Medenbach et al [29], which is 8.976Å^3 for Gd_2O_3, the density of the Gd_2O_3 films was calculated using the method described in chapter 4.5.1. The density was plotted versus both the deposition temperature and the number of ALD cycles and these can be seen in figures 5.24 and 5.25 respectively. The density of the Gd_2O_3 film is seen in figure 5.24 to decrease as wafer deposition temperature is increased. The decrease in film density is proposed to be due to a combination of two effects both related to an increased silicon uptake in the gadolinium oxide film with increasing deposition temperature. The first effect is a change of physical density which is caused by the variations in maximum packing density due to a change in the composition of the film. The second effect is due to a change in the average polarizability of the high-k films. This is

because the average polarizability of a material is highly dependent on the structural units and these will change with composition of a given film.

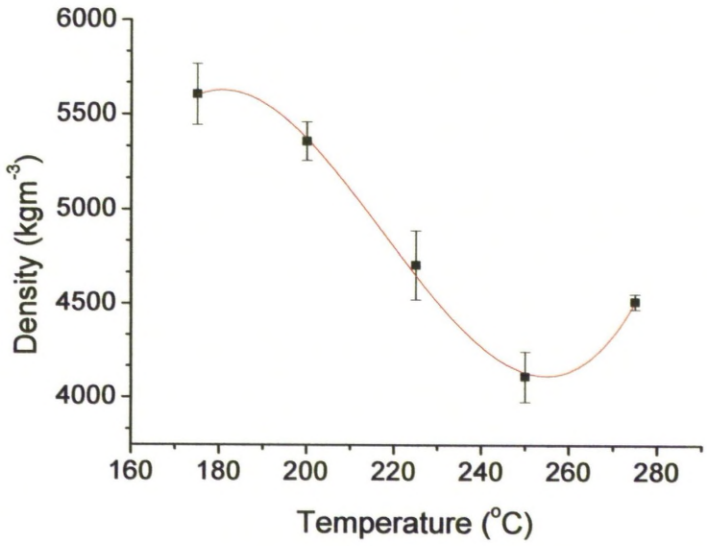

Figure 5.24- Plot showing the density of the ALD deposited films versus the wafer deposition temperature (100 ALD cycles)

Figure 5.25 shows that increasing the number of ALD cycles produces films of greater density. This is to be expected because increasing the number of deposition cycles will produce a thicker film with closer physical properties to those of bulk material. It is also noticeable that the maximum density of the gadolinium oxide film, taken from sample exposed to 300 ALD cycles, is about 13% lower than bulk gadolinium oxide and this could be due a combination of three reasons. The first is that the samples deposited at 200°C should be amorphous and so the packing density will be noticeably lower than for bulk (crystalline) gadolinium oxide. The second is that at a wafer deposition temperature of 200°C, a significant level of silicon has been introduced into the film from the precursor and it is very likely that this would have lowered the packing density and hence the physical density of the gadolinium oxide film. The final reason for the relatively low extracted density value is related to the second in that the silicon introduced will alter the polarizability, creating an erroneous shift in extracted density. It becomes apparent from the study, that there is a need to better understand the chemistry of gadolinium oxide films

containing significant levels of silicon so as to allow better modelling of the structural units and in turn, obtain more accurate values of polarizability.

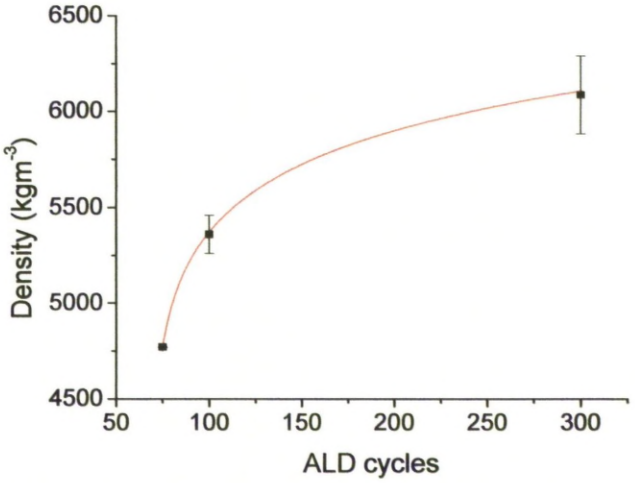

Figure 5.25 - Plot showing the density of the ALD deposited gadolinium oxide films versus the number of ALD cycles the sample is exposed to for a fixed wafer deposition temperature (200°C)

For samples deposited at 200°C, the band gap was extracted by from a Tauc plot; the values obtained are ~4.96 eV for 100 cycles and ~4.23 eV for 300 cycles, as shown in figure 5.26. It can be seen that increasing the number of ALD cycles results in a reduction of the effective band-gap, presumably because the defects at the band edge now extend further into the gap of the gadolinium oxide. The same procedure was carried out for the MBE Gd_2O_3 layers grown at 500°C and the results are included on figure 5.26. The ALD gadolinium oxide with only 100 ALD cycles is seen to have a similar effective band-gap to that of the MBE samples; however this is much lower than results reported elsewhere, of 6.4eV [30] and 5.3eV [20]. This disparity could be related to a defect related peak occurring close to the band edge of these gadolinium oxide films and that the actual band edge is higher in energy than that measured. A considerable absorption peak at 3eV in the measured data for the MBE deposited sample is evident on figure 5.26 and this could be due to a large defect in the band-gap of this film. The existence of this defect for the MBE gadolinium oxide films but not for the ALD gadolinium oxide films suggests that the ALD process is a better choice for the deposition of gadolinium oxide.

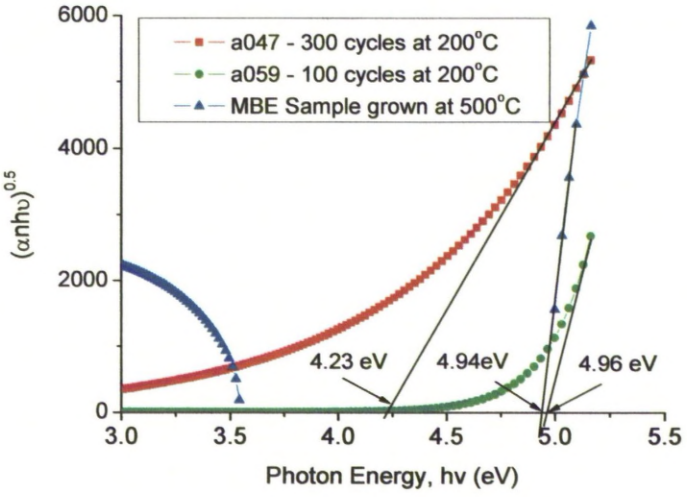

Figure 5.26 – Tauc plot showing (αhvn)1/2 versus photon energy for ALD Gd2O3 films with varying number of deposition cycles at 200°C and for an MBE sample grown at 500°C

The result of increasing the deposition temperature whilst keeping the number of cycles constant at 100 can be seen in figure 5.27. The absorption in the films is seen to increase, starting with the near band edge region at ~4eV. An abrupt change in absorption for samples which had different deposition or annealing conditions usually indicates a change in structure, stoichiometry or composition of the film. It is highly likely that the change is due to a change in structure as there is published evidence that a polycrystalline structure can be induced in some samples for temperatures > 250°C [31]. However, this precursor is known to introduce silicon into the gadolinium oxide layer and this would change the chemical composition of the film and possibly the structure, resulting in a change to the absorption spectra for the film. However the silicon concentrations shown in figure 5.21 do not exhibit an abrupt change between 225°C and 250°C, rather there is a gradual change. It could be assumed that such a change would produce a gradual change in absorption spectra with increasing temperature and not the sudden change seen at about 250°C. A structural change to the gadolinium oxide films could also explain the increase in the extracted density of the films for a deposition temperature of above 250°C in figure 5.24 earlier. This is because any formation of crystalline regions would result in an effective density closer to that of bulk gadolinium oxide

material. Therefore, an increase in density caused by the possible formation of nano-crystals could produce a higher density value than the decrease caused by the extra silicon incorporated into the layer.

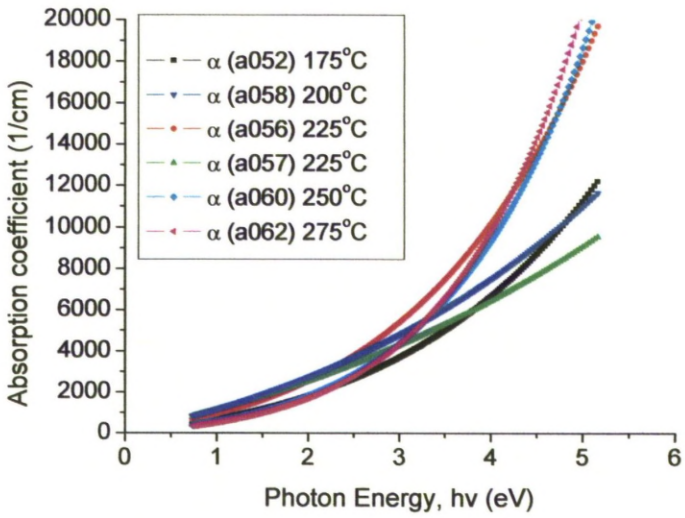

Figure 5.27 - Plot showing absorption coefficient versus photon energy for varying deposition temperature for a fixed number of deposition cycles (100 cycles)

5.4.2. Gadolinium silicate produced by high temperature annealing

Gadolinium oxide samples were deposited on n-type (100) wafers by liquid injection atomic layer deposition (LI-ALD), using a modified Aixtron 200FE AVD reactor. The precursor used here was $Gd[MeCp]_3$ in toluene solution, with H_2O as the oxidant, different to that used to deposit the previous gadolinium oxide films. The reason for the choice of precursor is to avoid introduction of silicon into the gadolinium oxide during deposition so that the effects of high temperature annealing on a pure gadolinium oxide film can be studied. It was known from diffusion experiments on e-beam deposited gadolinium oxide described earlier in this chapter, that high temperature annealing causes diffusion of silicon into the high-k film, improving both the CET and leakage. Therefore as e-beam methods are not compatible with the CMOS process, it was decided to see if the ALD process could produce similar results. The growth temperature for the samples was kept fixed at 250°C and film thickness was varied between 3 and 25nm by altering the number of ALD cycles. The gadolinium oxide layers were grown on 1.5nm native silicon dioxide; no

deliberate oxidation was employed. Some samples received an RTA in a nitrogen environment at temperatures between 800°C and 1050°C for 1s. The remaining samples were annealed at 900°C in nitrogen for time durations of between 1 and 100 s. Nitrogen was used to provide an inert atmosphere so as to avoid interfacial layer growth at the high temperatures. It is also typically one of the gases used by device manufacturers during high temperature drain activation annealing.

Medium energy ion scattering (MEIS) was carried out at the STFC Daresbury Laboratory facility using the same measurement and analysis technique as for the deposited silicate gadolinium oxide samples analysed earlier in this chapter. MEIS depth profiles for Gd, Si and O are given in figure 5.28 for a 5.5 nm thick sample for the as-deposited condition (a), and following RTA for 1s at 850 °C (b), 950 °C (c), and 1050 °C (d) in a nitrogen ambient. The MEIS data for the as-deposited sample reveal a thickness of 5.5 nm for the Gd_2O_3 layer on top of the native SiO_2 layer. When energy straggling and system resolution are taken into account, no Si is seen to be contained within the as-deposited gadolinium oxide layer as expected using this precursor and such low deposition temperatures. However silicon migration into the gadolinium oxide layer can be seen for the samples exposed to rapid thermal annealing above 850°C. Annealing at 850 °C for 1s can be seen to introduce only a small amount of Si into the layer towards the high-k and silicon interface. After annealing for 1s at 950 °C in nitrogen, the Si content in the layer is raised further but it can be seen that the silicon atoms have not reached the gadolinium oxide films surface yet and so a complete silicate layer has not yet been formed. Following 1050 °C annealing for 1s in a nitrogen environment it can be seen that the silicon has been diffused throughout the layer and that it has formed a complete silicate layer; $Gd_{2-x}Si_xO_{3-\delta}$, with a composition of x=1.

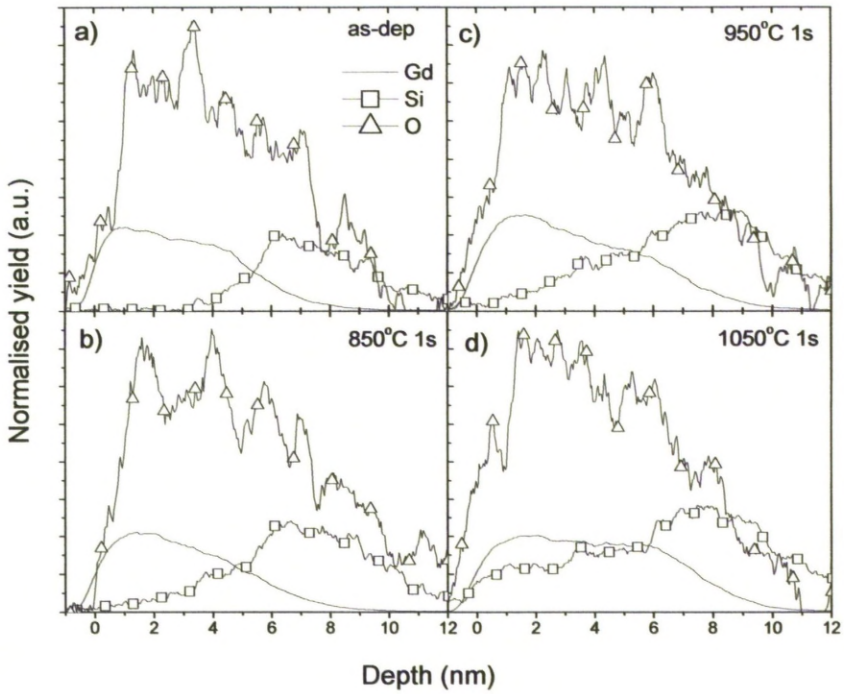

Figure 5.28 - MEIS depth profiles for ALD Gd_2O_3 films deposited on native oxide which have undergone RTA in nitrogen for 1s at different annealing temperatures

The reduced height of the Gd peak and greater width at 1050 °C, suggests that the original Gd_2O_3 and the original SiO_2 layer are completely intermixed. If the amount of Si in the original SiO_2 layer is considered then it is clear from the high Si content in the high-k layer that that some of the Si in the gadolinium silicate layer had originated from the substrate. The samples annealed at 900 °C for increasing duration has shown a similar result where annealing for time periods up to 100s in a nitrogen ambient produced a fully silicated high-k layer. The Si content of the layers annealed for 1s at different annealing temperatures in a nitrogen environment, was extracted using the same technique used in the previous ALD gadolinium oxide layers and is plotted in figure 5.29. Figure 5.30 shows that the silicon uptake follows an Arrhenius relationship and so the uptake of silicon is seen to be a thermally driven process. Comparison was made with TEM micrographs shown later in figure 5.32 and it was concluded

that some of the silicon dioxide interfacial layer may have been consumed.

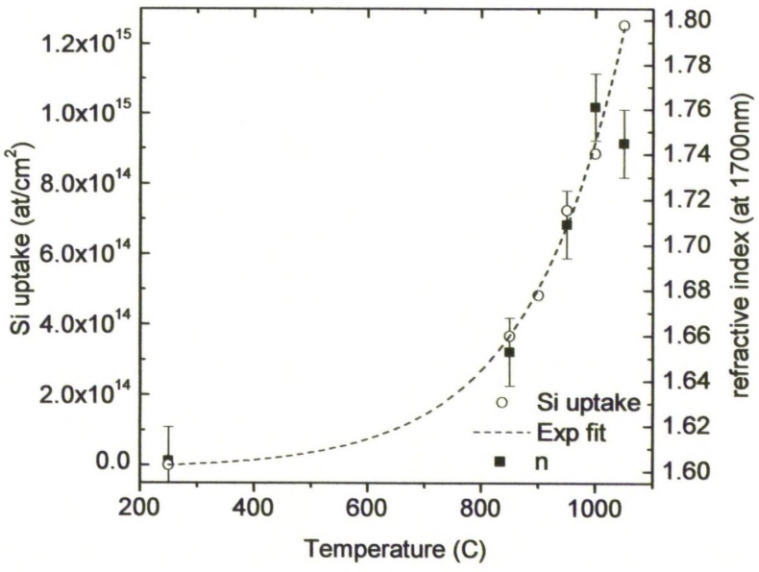

Figure 5.29 - Plot showing the average amount of silicon diffused into the films with annealing temperature. The change in refractive index for the samples at a wavelength of 1700nm is also given

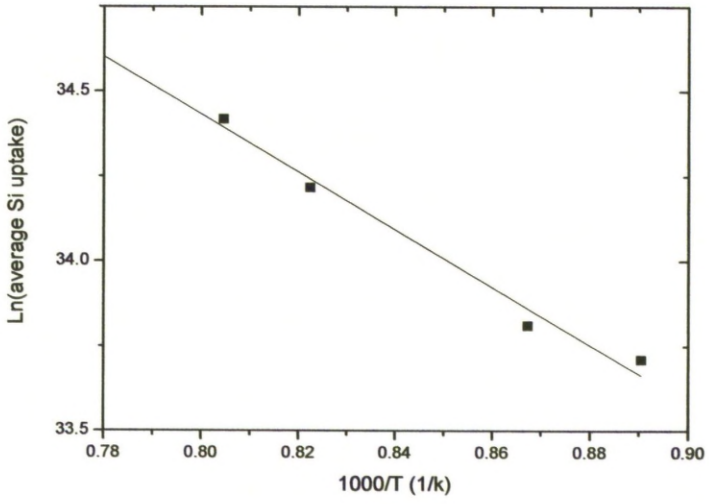

Figure 5.30 – Arrhenius plot of the average silicon uptake demonstrating that silicon diffusion is a thermally driven process

Ellipsometry data for all of the as-deposited and annealed gadolinium oxide films were measured using a J. A. Woollam M2000UI spectroscopic ellipsometer over a spectral range of 240-1700nm and for the incident angle

range of 65-80° in 5° steps. A sample of the silicon wafers was also measured before gadolinium oxide deposition to assess the thickness of the native oxide layer. For simplicity, this interfacial region was modelled in the WVASE32 software as an ideal silicon dioxide layer using a refractive index value from the database because the independent extraction of optical constants introduces significant difficulties for such thin films. The thickness was extracted by fitting over the whole measured spectral range. The thicknesses for native oxides were seen to be between 1.6-1.8nm. All gadolinium oxide films were modelled using the WVASE software comprising of:

a) silicon substrate,

b) fixed 1.6nm ideal silicon dioxide interfacial layer and

c) a Cauchy layer with an Urbach absorption tail to model the high-k layer.

The procedure was to fit the experimental data for thickness using the first Cauchy parameter within the transparent region (~600-1700nm) and then on subsequent fits, attempts at including either one or both the other Cauchy parameters was undertaken to see if this improved the model fit. The wavelength range was then expanded to the whole measured range and the model was fit to the experimental data including the Urbach tail to model near band edge absorptions. The optical constants at the longest measured wavelength were taken from the model and can be seen in figure 5.29.

There are two possible explanations for the increase in refractive index with annealing temperature. The first relates to crystallisation of the amorphous high-k layer. Increasing crystallisation of the high-k layer would produce a commensurate increase in refractive index. It can be seen from the XRD measurements in figure 5.31 that the level of crystallinity increases with annealing temperature even in the sample with the lowest annealing temperature. However the film is seen to reduce in crystallinity at annealing temperatures of above 1000°C and the refractive index value does not decrease until beyond 1000°C. It can also be seen that the refractive index does not decrease back to the value of the amorphous, as-deposited samples after annealing at 1050°C. However, the high temperature RTA could have removed porosity within the layer and so could be much denser than the as-deposited

layers, which would result in a higher refractive index. Another possibility is that the silicon diffused into the film serves to increase the refractive index over the measured spectral range. This at first seems unlikely as silicon inclusion in hafnium silicate films results in a decrease of the k-value of the film. However, the major influence on the k-value of films such as gadolinium oxide and hafnium oxide comes from lattice contributions which occur at a much longer wavelength. Therefore it is possible that a gadolinium silicate film could have a relatively high relative permittivity in this measured spectral range but have a lower static dielectric constant overall due to lower lattice contributions.

X-ray diffraction (XRD) was used on some of the thicker films (~25nm) to study the changes in crystallinity in the high-k layer during the high temperature annealing, and the results are shown in figure 5.31. The as-deposited sample contains no diffraction features in the XRD results indicating that the film is amorphous in the as-deposited state, at least within the resolution of XRD. After an RTA at 800 °C for 1s in nitrogen ambient, the XRD spectrum contains a diffraction peak centred at 28.6°. This diffraction angle has been reported to be caused by the formation of cubic Gd_2O_3 [35]. Increasing the annealing temperature to 850 °C and then 900 °C produces successively larger diffraction peaks, indicative of an increase in the level of crystallinity in the films. Above an annealing temperature of 950 °C, the trend can be seen to be reversed, with the diffraction peaks starting to reduce in magnitude with increasing annealing temperature. At the highest annealing temperature of 1050°C in this study, XRD shows very few diffraction features suggesting that the film is mostly amorphous. A shift in the peak position towards a slightly higher angle occurs with increasing temperature, and is consistent with the incorporation of additional silicon.

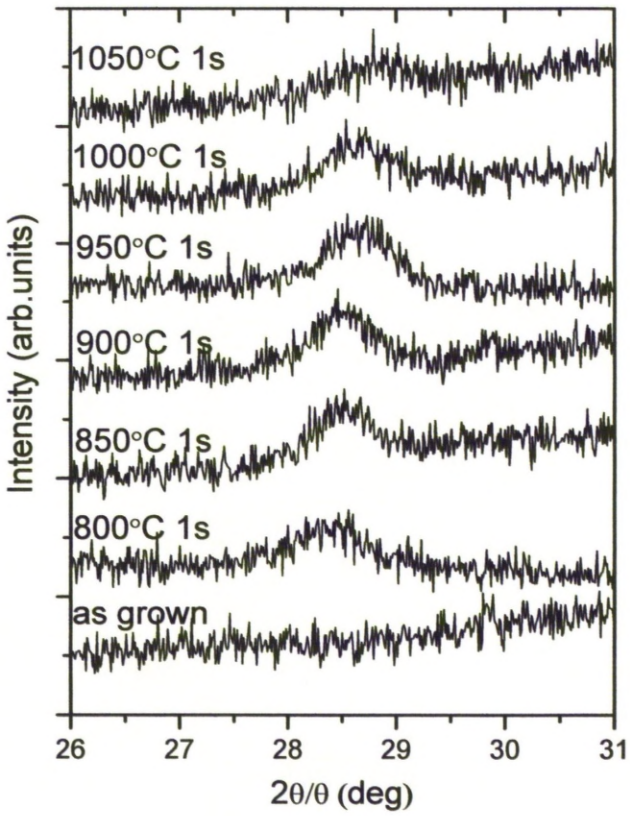

Figure 5.31 - X-ray diffraction (XRD) films of a gadolinium oxide film as deposited and annealed at temperatures between 800 and 1050C in nitrogen for 1s

When the MEIS and XRD results are compared, it becomes apparent that there are two competing mechanisms occurring during high temperature annealing. The first is the crystallization of the Gd_2O_3 layer into the cubic phase. The crystallization is temperature dependent as described before, with the level of crystallinity increasing with increasing temperature. Crystallization is the dominant mechanism at temperatures up to 900 °C. The second mechanism is the diffusion of Si into the high-k layer. This diffusion or silication, occurs at all temperatures studied, but is masked at low temperatures by the dominant crystallisation effects. As the annealing temperatures are increased however, a larger concentration of silicon is driven into the high-k layer and also this silicon is diffused further into the layer until it reaches the top interface. This is observed as a large reduction in the crystallinity of the high-k layer in XRD as the diffused silicon in the film converts the gadolinium oxide into an amorphous

silicate. Hence it is proposed that silicate formation is the dominant mechanism occurring during annealing at temperatures above 900 °C.

TEM images are displayed in figure 5.32 for a) a 5 nm thick as deposited film, and films annealed in a nitrogen ambient at b) 900 °C for 1s and c) 1000 °C for 1s. The as-deposited film can be seen to be essentially amorphous, although nano-crystallites have been observed in some TEM images. In the TEM micrograph of the film annealed at 900 °C (b), two distinct regions can be seen. The region, closest to the top surface of the high-k film, is composed of crystallites with an average width of approximately 7 nm. The narrow region next to the interfacial layer which contains the diffused silicon shows no long-range order and is amorphous. In the TEM image for the sample annealed at 1000 °C (c), it can be seen that the crystallites within the film annealed at 900°C have been removed, converting the high-k film to an amorphous layer. This again confirms that the incorporation of silicon into the film inhibits the formation of nano-crystallites. Grain boundaries are well known to be a source of leakage current pathways [32]. Therefore it would be expected that the amorphous film produced by the 1000 °C anneal should display better leakage current characteristics compared with the 900 °C sample, due to the removal of the crystalline regions and hence leakage current pathways at grain boundaries.

The reason why some crystallinity is observed in the sample annealed at 1000°C analysed by XRD which cannot be seen in the sample annealed at 1000°C analysed by TEM, can be explained by the films being different in thickness. The gadolinium oxide film analysed by TEM is approximately 5nm thick as-deposited and the film analysed by XRD is approximately 25nm thick as-deposited. A higher temperature anneal would be required to diffuse the silicon throughout the whole of the thicker XRD film in the same time as the 5nm TEM sample; or inversely, for the same annealing temperature, it would require a longer exposure time. This means that while silicon has reached the surface of the 5nm film analysed by TEM in the 1s exposure to 1000°C, it has not yet reached the surface of the thicker, XRD analysed sample. Hence the TEM sample is expected to be fully converted to an amorphous layer whereas the XRD analysed sample would still exhibit crystallites in the gadolinium oxide film above the silicate region.

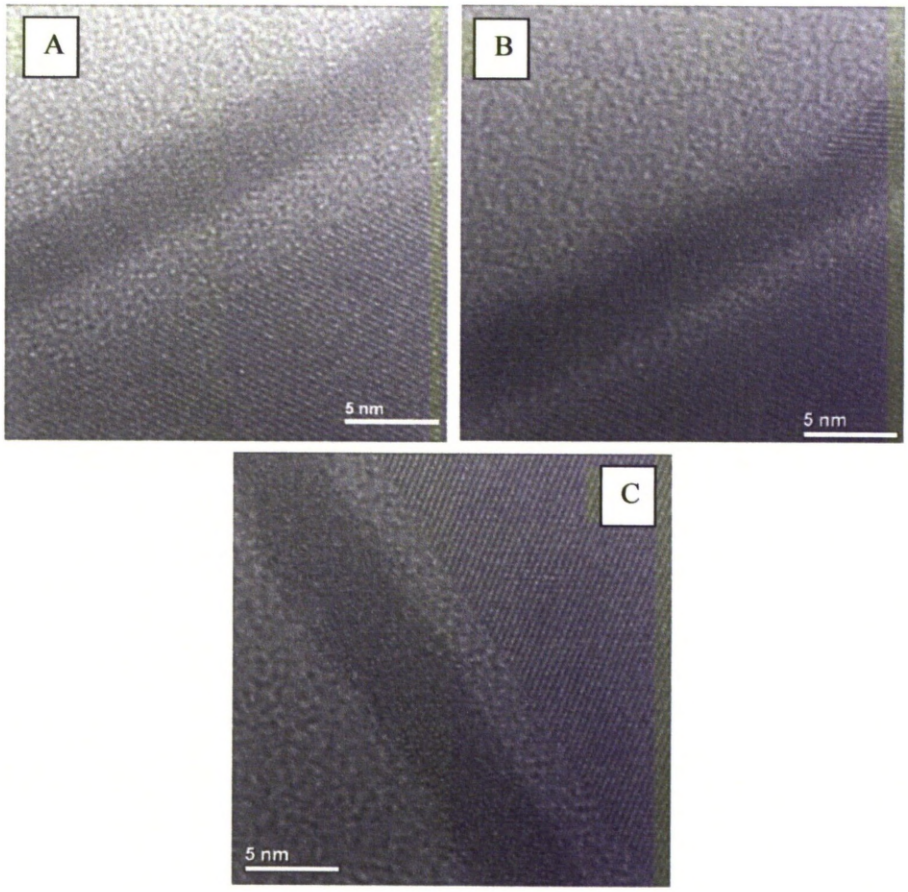

Figure 5.32 – TEM images for a) a 5nm thick as-deposited film, b) a film annealed at 900°C in N_2 for 1s and c) a film annealed at 1000°C in N_2 for 1s

The 'growth' rate of the silicate layer through the pure gadolinium oxide layer was investigated from the TEM images of thicker gadolinium oxide films which were exposed to various temperatures for 1s in nitrogen. The relationship between the thickness of the silicate layer and annealing temperature can be seen in figure 5.33.

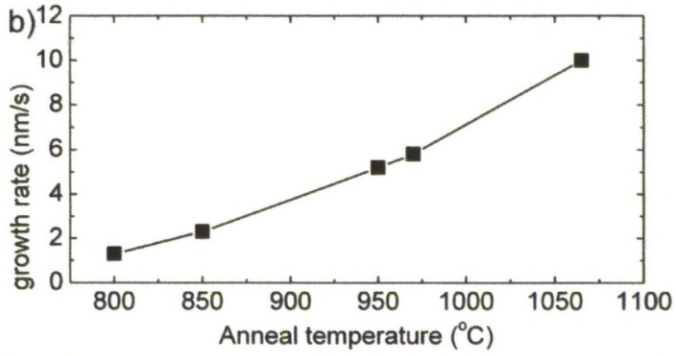

Figure 5.33 - Graph showing the diffusion rate of silicon when a gadolinium oxide is annealed at various temperatures in a nitrogen environment, for 1s

Contacts with an area of 4.5×10^{-4} cm^2 were formed by evaporating gold onto the samples through a shadow mask to form metal oxide semiconductor capacitor (MOSCAP) structures. Aluminium was deposited onto the cleaned, back side of the silicon substrate to form an ohmic contact. Capacitance-voltage measurements were carried out on all of the samples for a range of frequencies between 1 kHz and 1 MHz. A C-V characteristic measured at 1kHz for a sample with an initial 5.1nm gadolinium oxide layer deposited on a native silicon dioxide (from TEM) is shown in figure 5.34 for the as-deposited state and after annealing in nitrogen for 1s at either 900 or 1000°C. It can be seen in the C-V characteristic that there is a feature between 0-0.5V for the annealed samples which is not present for the as-deposited sample. This feature was frequency dependent and is likely to be caused by interface states; that is, the well-known P_{bo} centre related to dangling bonds at the silicon interface. It has been reported that interfaces of hi-k stacks incorporating an interfacial layer, show similar properties to those of pure SiO_2 [33]. As the feature is not present in the as-deposited sample, it shows that the native oxide provides a terminating surface for the silicon film; however modifications in this interfacial layer as a result of the annealing procedure degrade the terminating effect of the interfacial layer. If the effect is due to dangling bonds then a standard forming gas anneal should suppress it as has been seen in e-beam deposited gadolinium silicate layers [33]. There is also a positive flat-band shift on all of the samples which reduces significantly with annealing compared to the as-deposited sample. The CET values for the stack were calculated from the accumulation capacitance at 1 kHz and were 2.6 nm, 2.6 nm and 2.3 nm for the as-deposited and 900°C and

1000°C annealed samples respectively. For the 1000°C annealed sample the dielectric constant was extracted by assuming a two-layer model for the gate stack consisting of an interfacial layer of 1.6 nm obtained from TEM, with a dielectric constant of 3.9 assumed. This analysis yielded a k-value of 16 for the 4.6 nm silicate film.

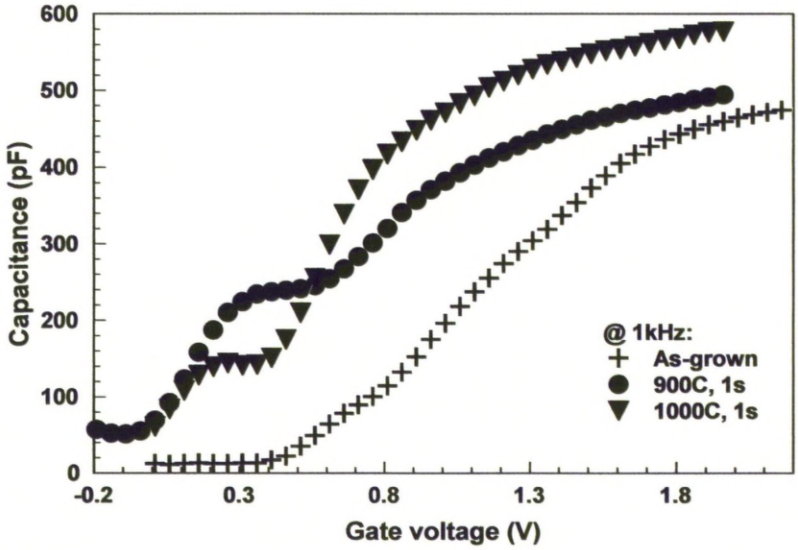

Figure 5.34 - Capacitance voltage measurement for a initially 5.1nm gadolinium oxide layer deposited on native oxide using ALD in the as-deposited state and after annealing in nitrogen for 1s at 900°C or 1000°C

Current-voltage (IV) measurements were performed on all of the samples and the IV measurements for the 5.1nm gadolinium oxide sample on native oxide are shown below in figure 5.35. The as-deposited layer appears to have the lowest leakage on the IV plot; however the large positive voltage (~0.5V) shift has to be taken into account. Comparing leakages around a bias voltage of 1.5V with the flat-band voltage value included, then the leakage for the sample annealed at 900C is still over an order of magnitude larger than the as-deposited sample, confirming that the anneal causes crystallisation of the gadolinium oxide layer introducing grain boundaries into the film which can act as conduction paths. The leakage after the 1000°C anneal is reduced significantly, by approximately an order of magnitude, compared to the 900°C annealed sample. Presumably the higher temperature anneal has converted the gadolinium oxide layer back to an amorphous layer as seen in the XRD and TEM results

and hence has removed the crystal grains which were acting as the conduction paths through the film.

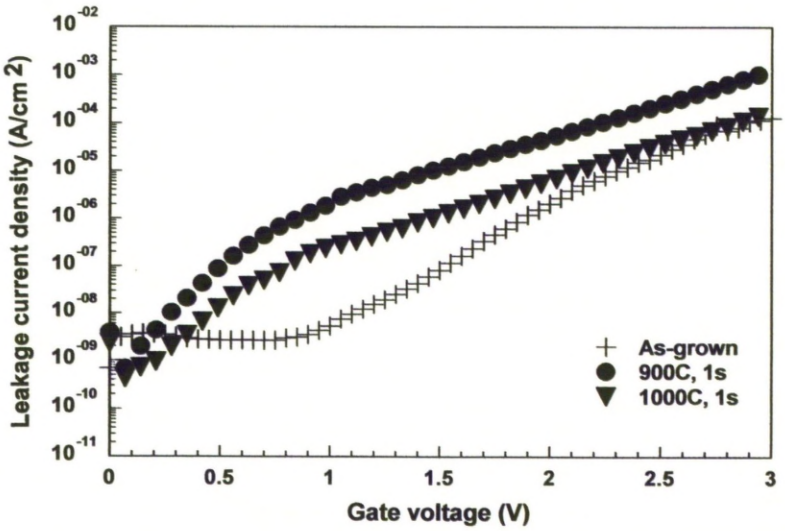

Figure 5.35 – J-V plot for an initial 5.1nm ALD Gd_2O_3 deposited on native oxide for the as-grown condition and after RTA in N_2 for 1s at 900°C and 1000°C

In figure 5.36 below, all of the extracted CET results from the ALD deposited films have been plotted versus the annealing temperature. The samples had been exposed to nitrogen for 1s. The samples labelled ALDN I are those deposited using the precursors $Gd[MeCp]_3$ and H_2O onto native oxide which have been analysed by physical methods in this section. It can be seen that annealing up to 900°C in nitrogen for one second significantly reduces the CET of the stack by causing densification of the high-k layer and consumption of a significant amount of the interfacial layer by the gadolinium oxide film during silicate formation. However, increasing the annealing temperature above 900°C does not produce any noticeable further reduction in CET. The initial 5.1nm gadolinium oxide sample analysed from the C-V plot of figure 5.36 is shown labelled as ALDN II. The reason why the CET after the same rapid thermal anneal is lower than that for the samples labelled ALDN I is that that this sample had a thinner gadolinium oxide layer, as-deposited; ALDN I samples were ~7nm. The samples labelled as ALDT I and ALDT II comprised ~6-10nm ALD gadolinium oxide films deposited on approximately 4nm and 2.5nm thermal oxides using the same precursor. With these samples, there was considerable consumption of the interfacial layer after annealing at 900°C in

nitrogen for 1s with the initial 4nm thermal oxide samples producing a stack with CET; similar to that of the gadolinium oxide films deposited on 1.6nm native oxide. After annealing in nitrogen for 1s at 1000°C, the CET slightly increases which is likely to be caused by interfacial oxide re-growth as seen earlier in this section. The samples with the 2.5nm thermal oxide IL, ALDT II, had a slightly lower CET than that of the ALD I samples and this is likely to be due to less of the interfacial layer remaining after the anneal.

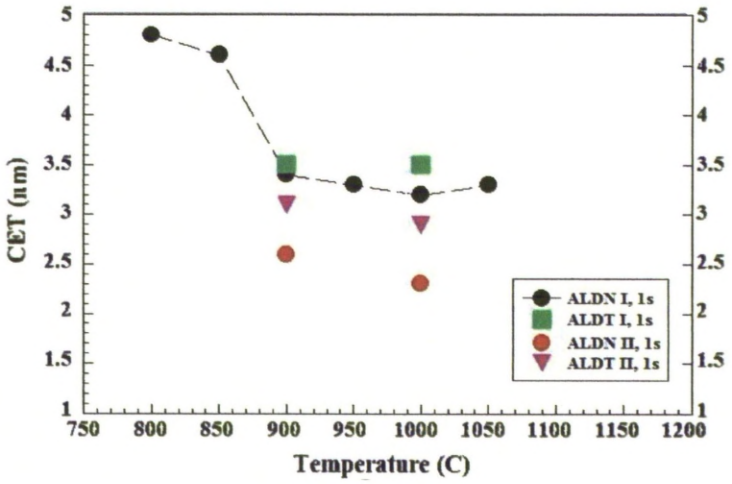

Figure 5.36 - Graph showing effect of RTA temperature on the CET of stacks containing ALD Gd_2O_3 films of different thickness deposited on thermal oxide (ALDT) or native oxide (ALDN)

5.5. Conclusions

Gadolinium silicate films can be produced either by incorporating silicon into a gadolinium oxide layer during deposition or by driving silicon into the layer using a high temperature RTA process, after deposition. It was observed in this chapter that for thicker interfacial layers, the CET of the gadolinium oxide stack reduced after RTA. However it was seen that there was a critical interfacial layer thickness where further thinning did not reduce CET and it was proposed that this was caused by residual oxygen within the chamber during RTA causing interfacial layer regrowth. The problem was solved by performing the RTA on a second set of samples with the vacuum valve left open. Using this process it was proven that the gadolinium silicate stacks produced from a e-beam evaporated gadolinium oxide could meet the ITRS requirements for the low

standby power 45nm node which was an objective within European project PULLNANO.

As ALD is seen as more industrially viable than e-beam evaporation, then attempts were made to use stacks deposited using this method meet the requirements for the 45nm node for low standby power. Incorporation of silicon into the gadolinium oxide layers was initially attempted by introducing it during deposition using the ALD precursors. Gadolinium silicate layers produced this way were studied in an attempt to optimise the deposition parameters and it was discovered that the best layers were produced for the lowest deposition temperature and for a low number of ALD cycles. As it had been found that high temperature annealing could produce a good quality gadolinium silicate film from an e-beam evaporated pure oxide then gadolinium oxide was deposited using ALD with a precursor not containing silicon. The RTA was then performed in nitrogen for a fixed time of one second at various annealing temperatures. The silicon content was observed to increase non-linearly with increasing annealing temperature. From TEM cross sections, a relationship between the thickness of gadolinium silicate formed with the annealing temperature was extracted to allow prediction of required annealing temperatures for future samples undergoing complete silication. Silication of a gadolinium oxide layer is seen to thermally stabilise the layer in an amorphous phase, which is advantageous for low leakage currents. The ALD gadolinium silicate stacks in this study did not meet the requirements for the 45nm technology node for low standby power, at least in this study. However this was expected to be due to the ALD gadolinium oxide layers being initially much thicker than the e-beam evaporated layers as $Gd[MeCp]_3$ was observed to have an uncertain growth rate and also the procedure of RTA with the vacuum valve open was not attempted for these layers. Optimisation of the deposition procedure should allow the ALD films to meet the stringent requirements for the 45nm node for low power.

5.6. References

[1] J. Kwo, M. Hong, A.R. Kortan, K.T. Queeney, Y.J. Cabal, J.P. Mannaerts et al, Appl. Phys. Lett., vol. 77, pp. 130-132, (2000)

[2] X. Wu, D. Landheer, G. I. Sproule, T. Quance, M. J. Graham, G. A. Botton, J. Vac. Sci. Technol. A, vol. 20, pp. 1141-1144, (2002)

[3] R. Lupták, K. Fröhlich, A. Rosová, K. Hušeková, M. Tapajna, D. Machajdík, M. Jergel, J.P. Espinós, C. Mansilla, Microelectronic Eng., vol. 80, pp. 154-157, (2005)

[4] M.P. Singh, C.S. Thakur, K. Shalini, S. Banerjee, N. Bhat, S.A. Shivashankar, J. Appl. Phys., vol. 96, pp. 5631-5637, (2004)

[5] B.A. Orlowski, E. Guziewicz, N.E. Orlowska, A. Bukowski, R.L. Johnson, Surf. Sci., vol. 507–510, pp. 218-222, (2002)

[6] P.Y. Kuei, C.C. Hu, App. Surf. Sci., vol. 254, pp. 5487-5491, (2008)

[7] H-H. Ko, L-B. Chang, M-J. Jeng, P-Y. Kuei, K-Y. Horng, Japanese J. Appl. Phys., vol. 44(5A), pp. 3205-3208, (2005)

[8] L.-Z. Hsieh, H.-H. Ko, P.-Y. Kuei, L.-B. Chang, M.-J. Jeng, J. Appl. Phys., vol. 98, 076110, (2005)

[9] Laha, H. J. Osten, A. Fissel, Appl. Phys Lett., vol. 90, 113508, (2007)

[10] M. Czernohorsky, E. Bugiel, H.J. Osten, A. Fissel, O. Kirfel, Appl. Phys. Lett., vol. 88, 152905, (2006)

[11] H.D.B. Gottlob, M. Schmidt, A. Stefani, M.C. Lemme, H. Kurz, I.Z. Mitrovic, W.M. Davey, S. Hall, M. Werner, P.R. Chalker, K. Cherkaoui, P.K. Hurley, J. Piscator, O. Engstrom, S.B. Newcomb, Microelectronic Eng., vol. 86, pp. 1642-1645, (2009)

[12] H.D.B. Gottlob, M. Schmidt, M.C. Lemme, H. Kurz, I.Z. Mitrovic, M. Werner, W.M. Davey, S. Hall, P.R. Chalker, K. Cherkaoui, P.K. Hurley, B. Raeissi, O. Engstrom, S.B. Newcomb, J. Vac. Sci. and Tech. B, vol. 27(1), pp. 249-252, (2009)

[13] A.C. Jones, H.C. Aspinall, P.R. Chalker, R.J. Potter, K. Kukli, A. Rahtu, M. Ritala, M. Leskala, Mater. Sci. and Eng. B, vol. 118, pp. 97-104, (2005)

[14] I.Z. Mitrovic, M. Werner, W.M. Davey, S. Hall, P.R. Chalker, H.D.B. Gottlob, M.C. Lemme, O. Engstrom, K. Cherkaoui, P.K. Hurley, 39th IEEE SISC 2008, San Diego, CA, USA, (2008)

[15] M. Werner, P.R. Chalker, W.M. Davey, I.Z. Mitrovic, S. Hall, I. Alexandrou, Appl. Phys. Lett. (2009) (submitted).

[16] J.W. Johnson, B. Luo, F. Ren, B. P. Gila, W. Krishnamoorthy et al, Appl. Phys. Lett., vol. 77, 3230-3232, (2000)

[17] M. Hong, J. Kwo, A. R. Kortan, J. P. Mannaerts, A. M. Sergent, Science, vol. 283, pp. 1897-1900, (1999)

[18] D. Jia, L. Lu, W.M. Yu, Opt. Commun., vol. 212, pp. 97-100, (2002)

[19] A. Fissel, M. Czernohorsky, H. J. Osten, J. Vac. Sci. Technol. B, vol. 24(4), pp. 2115-2118, (2006)

[20] D. P. Norton, Mater. Sci. and Eng., vol. 43, pp. 139-247, (2004)

[21] J. Robertson, Eur. J. Phys., vol. 28, 265-291, (2004)

[22] K.J. Hubbart, D.G. Schlom, J. Mater. Res., vol. 11, pp. 2757-2776, (1996)

[23] H.D.B. Gottlob, T. Echtermeyer, M. Schmidt, T. Mollenhauer, J.K. Efavi, T. Wahlbrink et al., IEEE Electron Dev. Lett., vol. 27, pp. 814-816, (2006)

[24] M. Czernohorsky, D. Tetzlaff, E. Bugiel, R. Dargis, H. J. Osten, H. D. B. Gottlob, M. Schmidt, M. C. Lemme, H Kurz, Semicond. Sci. Technol., vol. 23, 035010, (2008)

[25] H.D.B. Gottlob, M. Schmidt, A. Stefani, M.C. Lemme, H. Kurz, I.Z. Mitrovic, W.M. Davey, S. Hall, M. Werner, P.R. Chalker, K. Cherkaoui, P.K. Hurley, J. Piscator, O. Engström, S.B. Newcomb, Microelectronic Eng., vol. 86, pp. 1642-1645 (2009)

[26] P. Bailey, T.C.Q. Noakes and D.P. Woodruff, Surf. Sci., vol. 426, pp. 358-372, (1999)

[27] CRC Handbook of Chemistry and Physics, 84th Edition, CRC press, (2003)

[28] SIMNRA Program, M. Mayer, Max Planck Institute of Physics, http://www.rzg.mpg.de/~mam/

[29] O. Medenbach, D. Dettmar, R. D. Shannon, R. X. Fischer and W.M.Yen, J. Opt. A: Pure Appl. Opt., vol. 3, pp.174-177, (2001)

[30] T. Hattori, T. Yoshida, T. Shiraishi, K. Takahashi, H. Nohira, S. Joumori, et al., Microelectron Eng 72, pp. 283–287, (2004)

[31] O. Engström, B. Raeissi, S. Hall, O. Buiu, M.C. Lemme, H.D.B. Gottlob, P.K. Hurley and K. Cherkaoui, Solid-State Electron, vol. 51, pp. 622-626, (2007)

[32] K.P. McKenna and A.L. Shluger, Microelectronic Eng., vol. 86, pp. 1751-1755, (2009)

[33] P.K.Hurley, K.Cherkaoui, E.O'Connor, M.C.Lemme, H. D.B. Gottlob, M.Schmidt, S.Hall, Y.Lu, O.Buiu, B.Raeissi, J. Piscator and O.Engstrom, J. of Electrochem. Soc., vol. 155 pp.G13-G20 (2008)

[34] M. Hong, J. Kwo, S. N. G. Chu, J. P. Mannaerts, A. R. Kortan, H. M. Ng et al., J. Vac. Sci. Tech. B, vol. 20, pp. 1274-1277, 2002

[35] ICDS Card Nos. 86 – 2477 and 88 – 2165.

[36] J.A. Gupta, D. Landheer, J.P. Mcaffrey, and G.I. Sproule, Appl. Phys. Lett., vol. 78, pp. 1718, (2001)

Chapter 6

Summary and future work

6. Summary and future work

The purpose of this chapter is to discuss the implications of the presented results, to provide direction on what further work should be done to expand on the discoveries and to suggest improvements to the techniques which have been used. This thesis has presented results for both hafnia and gadolinium based materials, which are proposed as replacements for silicon dioxide as the gate dielectric in traditional CMOS.

In the fourth chapter of the thesis hafnium silicates were investigated in order to gain understanding of the relationship between the composition and the physical and electrical properties of the dielectric layer. A review of the topic revealed that little data had been published for different compositions of hafnium silicates using similar deposition techniques. In this study a wide range of silicate compositions were deposited using MOCVD on chemical interfacial oxides or those prepared by rapid thermal oxidation. It was shown that increasing silicon content within a hafnium silicate causes a decrease of the dielectric constant of the layer. This decrease in the dielectric constant for hafnium silicate films was seen to be due to a combination of the decrease of both the high frequency permittivity and the lattice contributions, with the latter being the dominant component. It was also shown that the band-edge of the material is shifted to higher energies with increasing silicon content resulting in a layer with a larger band-gap. This latter relationship was, however, seen to be non-linear in nature with a significant percentage of silicon required to produce a significant increase in the band-gap. A shallow band of defect energy levels was observed near the conduction band and as expected from the literature, these levels went deeper into the band-gap as silicon content increased, until the film was predominantly composed of silicon dioxide, at which point it is proposed that hafnium oxide provides an energy level close to the band-edge. This observation from optical analysis was consistent with electrical results whereby a hafnium rich, silicate layer displays high capacitance density and also greater leakage current density. These properties imply a trade-off is required to produce an optimum material for gate stack usage. Since this research was carried out, it was reported that Intel has incorporated a hafnium based dielectric into its microprocessor devices for the 45nm technology node, proving its industrial viability

In chapter five a novel candidate for a replacement gate dielectric system, namely gadolinium oxide, was investigated as part of the European integrated project PULLNANO with the target to meet the ITRS requirements for the 45nm technology node for low standby power, which utilises a gate length of 24nm. The material was initially deposited by e-beam evaporation to allow an easy route to samples and to develop a process for the production of an optimal gate stack. High temperature annealing was observed to cause silication of the gadolinium oxide layer which served to improve its thermal stability. An etching study of ultrathin thermal oxides allowed an investigation into the effects of the initial interfacial layer thickness on the properties of the gate stack. It was found that for thicker interfacial thermal oxide that the CET of the gate stack decreased, commensurately with initial silicon dioxide interfacial layer thickness. However, it was observed that for thinner initial interfacial layers, that CET increased after high temperature annealing and it was suggested that this increase was due to the presence of residual oxygen in the growth chamber. The latter problem was resolved by leaving the vacuum line of the rapid thermal processor open during annealing. This remedy allowed the realisation of prototype e-beam gadolinium silicate layers which met a CET of 1.3 nm and leakage current of < 120 mA/cm^2 ITRS requirements for low standby power application and hence allowed the academic cluster to exceed its targets for the PULLNANO project. The advantage to this technique is that it exploits the natural mixing of the interfacial and high-k layers which would occur during source/drain implant activation annealing in the standard CMOS process flow. The work then turned to the production of similar gadolinium silicate gate stacks using the more industrially viable ALD deposition process. Incorporation of silicon into the ALD gadolinium oxide layers during deposition was initially investigated and the process optimised using the measurement tools spectroscopic ellipsometry and MEIS. This procedure is feasible for incorporation of silicon into the high-k layer, but subsequent high temperature annealing during the CMOS production process would result in further incorporation of silicon into the layer. The silicon content could therefore become excessive. Due to the success and merits for the method of incorporation of silicon into the e-beam evaporated gadolinium oxide stacks during high temperature annealing, then pure ALD gadolinium oxide layers

were deposited and exposed to rapid thermal processing. From analysis of these layers, it was observed from XRD that ALD gadolinium oxide layers crystallise at temperatures below 800°C, which is lower than those used in traditional processing. However, the incorporation of silicon into the layer causes the material to convert to a thermally stable, amorphous state. Within the time constraints of the European project the ALD gadolinium silicate layers did not meet the ITRS requirements for the 45 nm technology node (24nm gate length) for low standby power; however, this is not expected to be caused by any intrinsic problems with the ALD process. In fact, the higher CET values of the ALD gadolinium silicate stacks was the result of a thicker than desired, initial gadolinium oxide layer which was due to growth rate instability problems with the precursor used. Further work is required to develop and calibrate the deposition step. The process used to avoid interfacial silicon dioxide layer re-growth in the e-beam samples was discovered too late to be applied to the ALD samples and so it is likely that the ALD gadolinium silicate stacks suffered significant CET reduction through this effect.

The research reported in the thesis has also made contributions in the use of spectroscopic ellipsometry as an analytical technique for these ultra-thin, sub-10nm dielectric layers. In the fourth chapter it was demonstrated that thickness values for 2.5 nm dielectric layers could be extracted using spectroscopic ellipsometry. This was proven by the agreement within fractions of a nanometre, for thickness values obtained using HRTEM and MEIS. Also in chapter four a novel spectroscopic ellipsometric procedure for estimation of high-k density was demonstrated. The technique was used to estimate the density of experimental hafnium silicate layers of hafnium molecular composition, x, between 0.3 and 1.0 and the results were shown to be within of 15% agreement with that extracted using XRR. The power of the technique was also shown in the fifth chapter where it was used in the optimisation of the deposition of gadolinium oxide layers with a precursor containing silicon. The technique showed that the density of the deposited gadolinium silicate reduced with increasing wafer deposition temperature and also identified an increase in density due to the expected formation of polycrystalline regions within the layers at the highest deposition temperatures. Finally, a method was presented to

allow delineation of the contributions to the dielectric function of lattice and electronic components, using CV and SE measurements. This technique should be valuable when screening new high-k systems.

The research within this thesis however, is by no means complete as there is still much to be investigated. For the gadolinium silicate layers, the logical first step would to be to investigate the current mechanisms occurring within both the e-beam and ALD films. This could provide insight into possible defects and hence allow further optimisation of the layers produced by both deposition techniques. Also work should be carried out to optimise further the thickness control in the deposition of pure gadolinium oxide films by ALD to allow the production of layers comparable in thickness to those produced by e-beam evaporation. This would allow stacks containing thin ALD gadolinium oxide to undergo the optimised RTA procedure as used for the e-beam samples to verify that ALD samples could meet the ITRS requirements for the 45nm technology node for low standby power. Another interesting line of research on the gadolinium silicate layers would be to identify the source of the silicon incorporated within the gadolinium oxide layers during high temperature annealing. At present it is currently unknown if the silicon is incorporated purely as a result of the consumption of the interfacial layer or if there is also a contribution from the silicon substrate.

A further avenue of research would be to improve the density estimation technique which was demonstrated within this thesis. The technique uses values for polarizability, which is a property of the structural units occurring within crystalline materials. As the accuracy of the technique has only been verified for one amorphous material, namely hafnium silicate, then it must be further confirmed by analysing other amorphous layers to ascertain the applicability of the approach to non-crystalline materials. Another possible improvement would be to investigate the use of more complex optical models to extract the density of a film from optical data as the physical basis for the Clausius-Mosotti equation is somewhat simplistic. Also, many of published polarizabilities of materials are derived through calculation and extrapolation of only a few characterised materials, drawing on trends within periodic table groups. Work

should be undertaken to measure the polarizability of such substances where accurate values are not yet known.

List of publications

M. Werner, P.R. Chalker, W.M. Davey, I.Z. Mitrovic, S. Hall and Y. Alexandrou

'Formation of Gadolinium silicate high-k gate dielectrics via a silicon drive-in mechanism'

To be submitted to J. Appl. Phys. (2009)

W.M. Davey, O. Buiu, M. Werner, I.Z. Mitrovic, S. Hall and P. Chalker

'Estimate of dielectric density using spectroscopic ellipsometry'

Microelectronic Eng., Vol. 86, pp. 1905-1907, (2009)

H.D.B. Gottlob, M. Schmidt, A. Stefani, M.C. Lemme, H. Kurz, I.Z. Mitrovic, W.M. Davey, S. Hall, M. Werner, P.R. Chalker, K. Cherkaoui, P.K. Hurley, J. Piscator, O. Engström and S.B. Newcomb

'Scaling potential and MOSFET integration of thermally stable Gd silicate dielectrics'

Microelectron. Eng., Vol. 86, pp. 1642-1645, (2009)

H. D. B. Gottlob, A. Stefani, M. Schmidt, M. C. Lemme, H. Kurz, I. Z. Mitrovic, M. Werner, W. M. Davey, S. Hall, P. R. Chalker, K. Cherkaoui, P. K. Hurley, J. Piscator, O. Engström and S. B. Newcomb

'Gd silicate: A high-k dielectric compatible with high temperature annealing'

J. Vac. Sci. and Tech. B, Vol. 27, pp. 352-355, (2009)

O. Buiu, W. Davey, Y. Lu, I.Z. Mitrovic and S. Hall

'Ellipsometric analysis of mixed metal oxides thin films.'
Thin Solid Films, vol. 517, pp. 453-455, (2008)

S. Bernardini, M. MacKenzie, O. Buiu, P. Bailey, T.C.Q. Noakes, W.M. Davey, B. Hamilton and S. Hall

Chemical and optical profiling of ultra thin high-k dielectrics on silicon'

Thin solid films, vol. 517, pp. 459-461 (2008)

H.D.B. Gottlob, M. Schmidt, M.C. Lemme, H. Kurz, I.Z. Mitrovic, M. Werner, W.M. Davey, S. Hall, P.R. Chalker, K. Cherkaoui, P.K. Hurley, B. Raeissi, O. Engström, and S.B. Newcomb

'Gd silicate: A High-k Dielectric Compatible with High Temperature Annealing.'

J. Vac. Sci. and Tech. B, Vol. 27 (1), pp. 249-252 (2009)

Y. Lu, S. Hall, I. Z. Mitrovic, W. M. Davey, B. Raeissi, O. Engstrom, K. Cherkaoui, S. Monaghan, P. K. Hurley, H. D. B. Gottlob, and M. C. Lemme

'Leakage current effects on C-V plots of high-k MOS capacitors.'

J. Vac. Sci. and Tech. B, Vol. 27, pp. 352-355 (2009)

I.Z. Mitrovic, O. Buiu, S. Hall, C. Bungey, T. Wagner, W.M. Davey and Y. Lu

'Electrical and structural properties of ALD hafnium silicate thin films.'

Microelectron. Rel., Vol. 47, pp.645-648 (2007)

Y. Lu, O. Buiu, S. Hall, I.Z. Mitrovic, W. Davey, R. J. Potter and P.R. Chalker

'Tuneable electrical properties of hafnium aluminate gate dielectrics deposited by metal organic chemical vapour deposition'

Microelectron. Rel., Vol. 47, pp. 722-725 (2007)

S. Hall, O. Buiu, I. Mitrovic, Y. Lu and W. M. Davey

'High-k dielectric stacks for nanoscaled SOI devices.' (invited),

Proc. NATO ARW, pp. 33 – 58, Springer, (2007)

Conference presentations

W.M. Davey, O. Buiu, M. Werner, I.Z. Mitrovic, S. Hall and P. Chalker

'Estimate of dielectric density using spectroscopic ellipsometry'

16th Conference of Insulating Films on Semiconductors, July 2009

(Oral presentation)

O. Buiu, M. Werner, W. M. Davey, Y. Lu, S. Hall and P. Chalker

'Optimization of low temperature ALD gadolinium oxide films for gate dielectric stack applications.' 4th International Conference in Spectroscopic Ellipsometry, June 2007

(Oral presentation)

O. Buiu, W. Davey, Y. Lu, I.Z. Mitrovic and S. Hall

'Ellipsometric analysis of mixed metal oxides thin films.'
5th International Conference on Silicon Epitaxy and Hetrostructures, May 2007

(Poster)

O. Buiu, S. Hall and W. Davey

'Multi – angle spectroscopic ellipsometry investigation of hafnium silicates thin films'

4th Workshop Ellipsometry, February 2006,

(Poster)